

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 23:43:49 ; Search time 914 Seconds
(without alignments)
10183.591 Million cell updates/sec

Title: US-09-931-733-1
Perfect score: 2191
Sequence: 1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04.*
1: Geneseq1980s.*
2: Geneseq1990s.*
3: Geneseq2000s.*
4: Geneseq2001as.*
5: Geneseq2001bs.*
6: Geneseq2002s.*
7: Geneseq2003as.*
8: Geneseq2003bs.*
9: Geneseq2003cs.*
10: Geneseq2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	1381.8	63.1	9742	6	ABL70480 Chemical
2	1330.6	60.7	9742	6	ABL70479 Chemical
3	240.2	11.0	2775	4	Aah72785 Human cer
4	240	11.0	2709	6	Abk84059 Human CDN
5	240	11.0	2709	7	Abx10429 DNA encod
6	240	11.0	2709	9	Acf79937 Breast Ca
7	239.4	10.9	504	3	Aac00830 Human sec
8	238.4	10.9	2716	6	Aas95004 Human DNA
9	230.2	10.5	640	3	Aac00831 Human sec
10	228.9	10.4	2736	6	Abq54618 Human ova
C 11	165.2	7.5	15732	6	Abk28234 DNA trans
C 12	165.2	7.5	15732	6	Aas4388 Chemical
13	157.4	7.2	15732	4	Abx62895 Human act
14	157.4	7.2	15732	6	Abk28233 Human CDN
15	157	7.2	2905	7	Aac00829 Human sec
16	157	7.2	2906	6	Abx52426 Human CDN
17	115.6	5.3	459	3	Aac00829 Human sec
18	106.4	4.9	1892	6	Aal5558 Human nuc
19	106.4	4.9	2817	7	Aak86425 Human imm
C 20	80.2	3.7	768	4	Aak86423 Human imm
C 21	80.2	3.7	768	4	Aak86426 Human imm
C 22	80.2	3.7	768	4	Aak86426 Human imm
C 23	79.4	3.6	175561	7	Aad55694 Human THB

24	77.6	3.5	6073	4	Aak82559 Human imm
25	77.6	3.5	6074	4	Aak82558 Human imm
C 26	77	3.5	7404	4	Aak70264 Human imm
C 27	77	3.5	7409	4	Aak70263 Human imm
C 28	76.8	3.5	1348	6	Aaf90649 Human TAN
C 29	76.8	3.5	1348	6	Aad44203 Human TAN
C 30	76.2	3.5	604	5	Abv58969 Human GPC
C 31	75.8	3.5	12932	9	Adc85672 Human GPC
C 32	74.4	3.4	84607	2	Aax90847 Human PAC
C 33	74.2	3.4	110000	6	Abn83429 Human tra
C 34	74.2	3.4	126512	6	Abq7530 Human ami
C 35	73.6	3.4	9377	6	Continuation (2 of
36	73.4	3.4	110000	7	Continuation (2 of
C 37	73.2	3.3	488	4	Aai37596 Probe #62
C 38	73.2	3.3	488	4	Aak1720 Human bon
C 39	73.2	3.3	488	4	Abx1404 Human liv
C 40	73.2	3.3	488	6	Abx06476 Human gen
41	73	3.3	300000	9	Abx06476 Human gen
42	72.8	3.3	136284	6	Abx83575 Human PTP
43	72.8	3.3	249999	7	Abx83575 Human CDN
44	72.2	3.3	700	4	Abz80229 Human tra
45	71.2	3.2	11481	4	Aah93186 Human inf
C 45	71.2	3.2	11481	4	Aak71590 Human imm

ALIGNMENTS

RESULT 1

ABL70480/C
ID ABL70480 standard; DNA; 9742 BP.

XX ABL70480;

DT 01-JUL-2002 (first entry)

XX Chemically treated cell signalling DNA sequence complementary to#185.

DE Cell signalling; cytosine methylation; cell signalling disease; cancer;
tumour; cytostatic; ds.

XX Unidentified.

XX WO200202807-A2.

XX 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP007471.

XX 30-JUN-2000; 2000DE-01032529.

PR 01-SEP-2000; 2000DE-01043826.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-154758/20.

XX Nucleic acid, useful for diagnosis and therapy of diseases associated
with cell signalling e.g. cancer, comprises chemically modified genomic
sequences of genes associated with cell signalling.

XX Claim 1; SEQ ID NO 370; 24pp + Sequence Listing; English.

XX The invention relates to a nucleic acid comprising a sequence of at least
18 bases of a segment of chemically pretreated DNA of genes associated
with cell signalling. The activity of the modified sequences of the
invention may be described as cytostatic. The object of the invention is
to provide the chemically modified DNA of genes associated with cell
signalling, as well as oligonucleotides and/or PNA-oligomers for
detecting cytosine methylations, as well as a method which is
particularly suitable for the diagnosis and/or therapy of genetic and
epigenetic parameters of genes associated with cell signalling. The
chemically modified DNA provided by the invention is useful for diagnosis

Mon Mar 29 09:55:00 2004

CC	and therapy of diseases such as solid tumours and cancer. The sequences	889	GATTAAACAGAGTCTCTCCTCAGCTACGCTACCGGAGTGTGAGTCTTCCAGGATTTTCAAGTCT	948
CC	given in records ABL70111-ABL70626 represent chemically pre-treated	5584	AAATTAAACAAACTCTCTCAGCTACCGGAGTGTGAGTCTTCCAGGATTTTCAAGTCT	5525
CC	genomic DNA's of genes associated with cell signalling. Note: The	949	TCGATGCTGTCATGTTGAGAGTCCACTATCCAGCTTTCTTTTATCCAGGATTAATGACAA	1008
CC	sequence data for this patent is not represented in the printed	5524	TCGATGCTGTCATGTTGAGAGTCCACTATCCAGCTTTCTTTTATCCAGGATTAATGACAA	5465
CC	specification, but is based on sequence information supplied by the	1009	ATCTGAGCATTCGTTGAATCTTAATCATTTAGGAGTAAAGTAACTAGTCTGTTCAAGGA	1068
CC	European Patent Office	5464	ATCTGAGCATTCGTTGAATCTTAATCATTTAGGAGTAAAGTAACTAGTCTGTTCAAGGA	5405
XX	Sequence 9742 BP; 2945 A; 88 C; 1819 G; 4890 T; 0 U; 0 Other;	1069	CAGCTAGTGTAAACTCTCTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1128
XX	Query Match 63.1%; Score 1381.8; DB 6; Length 9742;	5404	CAACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	5345
XX	Best Local Similarity 80.8%; Pred. No. 0;	1129	TTCTGATCTGGAAGCCGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1188
XX	Matches 1787; Conservative 0; Mismatches 402; Indels 24; Gaps 14;	5344	TTCTGATCTGGAAGCCGAGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	5285
QY	1 TACATAATGAGAGATTTTATTCAGATAATGCTGATTAATTAATTAATTAATTAATTAATTA	1189	TATTTAAACAAATAGGAGTTCATAGAAGACATTTCCCAACAGGAGATGAGTAAAGT	1248
DB	6482 TACATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	5284	TATTTAAACAAATAGGAGTTCATAGAAGACATTTCCCAACAGGAGATGAGTAAAGT	5225
QY	61 ATATGATGCTACATAA-TTAT	1249	ATTGCTAGCAATAGATGCTCTCTACCAAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT	1308
QY	6422 AT	5224	ATTGCTAGCAATAGATGCTCTCTACCAAGATCTCTCTCTCTCTCTCTCTCTCTCTCTCT	5165
QY	119 GGAATTTTGTATATCTAT	1309	GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1368
DB	6362 AAAAAATTTTAT	5164	AAAAAATTTTAT	5105
QY	6303 AAAAAATTTTAT	1369	ATAAGTCTCTAATGACTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG	1428
DB	238 GCGACTGCCACAAAGTGGGGAATGACATTTTAAAGCATTTCTTTGCTTCTATGATG	5045	ATAAGTCTCTAATGACTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG	5045
QY	6302 ACCACTACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	1429	CTTTTCACTATCGAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1487
DB	6243 AAAAAATTTTAT	5044	CTTTTCACTATCGAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	4985
QY	239 AAAAAATTTTAT	1488	GATCAAGTGTATTTTCTGTTTACATGATATATGATATATGATATATGATATATGATAT	1547
DB	6242 AAAAAATTTTAT	4984	ATATCAAAATATATTTTCTGTTTACATGATATATGATATATGATATATGATATATGAT	4925
QY	299 ACCGAGTTTGTGTGATGCTAGT-----CCCAATGAGACATTTACTGCAATATAAATG	1548	TCGTGACATCACCCGAAATTTTAACTGACATGATATGATATATGATATATGATATATG	1607
DB	6183 ACCGCAATTTAT	4924	TATACATCACCCGAAATTTTAACTGACATGATATGATATATGATATATGATATATG	4865
QY	353 AT	1608	TCCCTCTCACCATGCTCTCAGCCCTTCCAGGAGGAGTCTTCCACATTTGTTCTCTCTCT	1667
DB	6123 ATTAT	4864	TCCCTCTCACCATGCTCTCAGCCCTTCCAGGAGGAGTCTTCCACATTTGTTCTCTCTCT	4805
QY	413 CTTTGAACCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1668	ACGGAAGCTTAAACACATCTGTTGGGAGGAGTCTTCCACATTTGTTCTCTCTCTCTCT	1727
DB	6064 CTTTGAACCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	4804	ACGGAAGCTTAAACACATCTGTTGGGAGGAGTCTTCCACATTTGTTCTCTCTCTCTCT	4745
QY	473 TTAAGGCGCACTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	1728	T-GGGGAATAGCAGATAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1786
DB	6004 TTAAGGCGCACTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	4744	TAAACAAATTAACAAATTAACAAATTAACAAATTAACAAATTAACAAATTAACAAATTA	4685
QY	533 GAACACCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	1787	AGGCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	1846
DB	5944 GAACACCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG	4684	AAACCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	4625
QY	592 TCCCTTTTGTAT	1847	CAACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1906
DB	5884 TCCCTTTTGTAT	4624	CAACATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	4565
QY	652 GGCCCTAGTTTGGATTAACCAATCAGCTTTCTTCTGCTGATGATGATGATGATGATG	1907	AAGCAATGATTTCTGCAATTTCTGCAATTTCTGCAATTTCTGCAATTTCTGCAATTTCT	1966
DB	5824 AACCTTAATTTAAATTAACCAATCAGCTTTCTTCTGCTGATGATGATGATGATG	4564	AAACATTAATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	4505
QY	712 ACATATTAACAGTCTAT---GGGCCCTTTTCCACATGATGATGATGATGATGATGATG	1967	C-----AGGTGAGTACTGTTCTGTTTGTAAATATGATGATGATGATGATGATGATGATG	2021
DB	5764 ACATATTAACAGTCTAT---GGGCCCTTTTCCACATGATGATGATGATGATGATG			
QY	769 GCTGCTGGGTAAACCAATCATCTGCTAGTGTGATGATGATGATGATGATGATGATGATG			
DB	5704 ACTACTAAATTAACCAATCATCTGCTAGTGTGATGATGATGATGATGATGATGATG			
QY	829 TACAGCAGCTTAAGGGGACCCAGTGTCTTAAAGAGAGAGAGAGAGAGAGAGAGAGAG			
DB	5644 TACACACAGCTTAAGGGGACCCAGTGTCTTAAAGAGAGAGAGAGAGAGAGAGAGAG			

Qy	2140	GAAGATAACAGGCACTTTCTTA-TTCACCAAACTGAGGACTCAACACAGAGAAA	2191
Db	5419	GAAGATAACAGGCACTTTCTTA-TTCACCAAACTGAGGACTCAACACAGAGAAA	5471
Qy	AAH72785	standard; cDNA; 2775 BP.	
Db	AAH72785	(first entry)	
Qy	19-SEP-2001	Human cervical cancer marker nucleic acid 4059.	
Db	19-SEP-2001	Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.	
Qy	200000-142467-A2.	Homo sapiens.	
Db	200000-142467-A2.	WO200142467-A2.	
Qy	200000-0203791P.	14-JUN-2001.	
Db	200000-0203791P.	08-DEC-2000; 200000-US033312.	
Qy	200000-0203791P.	08-DEC-1999; 99US-0169681P.	
Db	200000-0203791P.	21-DEC-1999; 99US-0171350P.	
Qy	200000-0203791P.	14-NAR-2000; 2000US-0189315P.	
Db	200000-0203791P.	12-MAY-2000; 2000US-0203791P.	
Qy	200000-0203791P.	09-JUN-2000; 2000US-0210600P.	
Db	200000-0203791P.	21-JUL-2000; 2000US-0220114P.	
Qy	200000-0203791P.	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
Db	200000-0203791P.	Schlegel R, Deeds J, Berger A, Zhao X;	
Qy	200000-0203791P.	WPI; 2001-375006/39.	
Db	200000-0203791P.	New isolated nucleic acid for diagnosing and treating cervical cancer and for assessing and detecting compounds for treating the cancer.	
Qy	200000-0203791P.	Claim 1; Page 850-851; 1051pp; English.	
Db	200000-0203791P.	The invention relates to novel genes (AAH68727-AAH73383) associated with cervical cancer with cytostatic activity. The nucleic acids and encoded polypeptides are useful: to assess if a patient is afflicted with cervical cancer or has a pre-malignant condition; to monitor the progression of cervical cancer or a premalignant condition in a patient; and to select and/or assess the efficacy of a compound or therapy for inhibiting cervical cancer in a patient. The nucleic acids may also be useful for gene therapy	
Qy	200000-0203791P.	Sequence 2775 BP; 971 A; 559 C; 566 G; 664 T; 0 U; 15 Other;	
Db	200000-0203791P.	Query Match 11.0%; Score 240.2; DB 4; Length 2775;	
Qy	200000-0203791P.	Best Local Similarity 93.3%; Pred. No. 1.9e-50;	
Db	200000-0203791P.	Matches 251; Conservative 0; Mismatches 18; Indels 0; Gaps 0;	
Qy	1701	CCACATTTGTTTCTTCTCTCCATTTTCTCTGGGGAATAGCAGATAGGAGCAAGCCAGCT	1760
Db	10	CCGCGTCGACTACGGCTCGAGAAAGCCCGGGAATAGCAGATAGGAGCAAGCCAGCT	69
Qy	1761	AGTCAGTAACTAAGTCACTCAACCAAGGCTTTTCTTCTTGTATCTTTCAGATACCT	1820
Db	70	AGTCAGTAACTAAGTCACTCAACCAAGGCTTTTCTTGTATCTTTCAGATACCT	129
Qy	1821	CATTTTCTTAGCGTTTCTGGAGATTACACATCTCGGTTCCGTTCTGGGAACCTTTAC	1880
Db	130	CATTTTCTTAGCGTTTCTGGAGATTACACATCTCGGTTCCGTTCTGGGAACCTTTAC	189
Qy	1881	TGATTTATCTCCCTCCCTCACACAAATAGCATTTCTGAGATCTCAAG	1940
Db	190	TGATTTATCTCCCTCCCTCACACAAATAGCATTTCTGAGATCTCAAG	249

XX Human, 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW gene therapy; chromosome mapping; ss.
XX Homo sapiens.
XX EPI033401-A2.
XX PD 06-SEP-2000.
XX 21-FEB-2000; 2000EP-00200610.
XX 26-FEB-1999; 99US-0122487P.
XX (GSEST) GENSET.
XX Dumas Milne Edwards J, Duclert A, Giordano J;
XX WPI; 2000-500381/45.
XX P-FSDB; AAG00824.
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX Claim 1; SEQ ID NO 828; 71pp + Sequence Listing; English.
XX The present sequence is one of a large number of 5' ESTs derived from
CC mRNAs encoding secreted proteins. An ORF has been identified within the
CC sequence. The 5' ESTs were prepared from total human RNAs or poly(A)⁺ RNAs
CC derived from 30 different tissues. EST sequences usually correspond
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC well suited for isolating cDNA sequences derived from the 5' ends of
CC mRNAs and even in those cases where longer cDNA sequences have been
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC mRNAs with intact 5' ends and can therefore be used to obtain full length
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC gene therapy and chromosome mapping procedures. They are used to obtain
CC upstream regulatory sequences and to design expression and secretion
CC vectors
XX
SQ Sequence 504 BP; 163 A; 90 C; 97 G; 154 T; 0 U; 0 Other;
Query Match 10.9%; Score 239.4; DB 3; Length 504;
Best Local Similarity 99.6%; Pred. No. 1.5e-50;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1729 GGGGAATAGCAGAAATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAG 1788
Db 11 GGGCAATAGCAGAAATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAG 70
QY 1789 GCCCTTTTTCCTTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 1848
Db 71 GCCCTTTTTCCTTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 130
QY 1849 ACATCTCGGTTCCGTTTCTGGGAACCTTCTGATTTATCTCCCGCTCACACAAATAA 1908
Db 131 ACATCTCGGTTCCGTTTCTGGGAACCTTCTGATTTATCTCCCGCTCACACAAATAA 190
QY 1909 GCATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCA 1968
Db 191 GCATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCA 250
QY 1969 G 1969
Db 251 G 251
RESULT 8
AAS95004
ID AAS95004 standard; DNA; 2716 BP.
XX

AC AAS95004;
XX 14-FEB-2002 (first entry)
XX Human DNA sequence #259 expressed during foam cell differentiation.
XX Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX Homo sapiens.
XX WO200177389-A2.
XX 18-OCT-2001.
XX 04-APR-2001; 2001WO-US011128.
XX 05-APR-2000; 2000US-0195106P.
XX (INCY-) INCYTE GENOMICS INC.
XX Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GJ, Mikita T;
PI Tai J;
XX WPI; 2002-010925/01.
XX Composition useful for diagnosis of conditions, disorders or diseases
PT associated with atherosclerosis, comprises several polynucleotides that
PT are differentially expressed in foam cell development.
XX
PS Claim 1; Page 300-301; 315pp; English.
XX The present invention relates to the isolation of human polynucleotide
CC sequences that are differentially expressed during foam cell
CC differentiation. The polynucleotide sequences of the invention or a
CC composition comprising these polynucleotides are useful as a high
CC throughput method for detecting altered expression of one or more
CC polynucleotides in a sample. The polynucleotides can be used in the
CC diagnosis of disorders associated with foam cell development such as
CC atherosclerosis, cerebral stroke, and cardiovascular disorders such as
CC coronary artery disease. The polynucleotide sequences can also be used as
CC PCR primers and probes. The polynucleotides of the invention are also
CC useful in gene therapy. AAS94746-AAS95021 represent the human
CC polynucleotide sequences of the invention which are differentially
CC expressed during foam cell differentiation
XX
SQ Sequence 2716 BP; 964 A; 537 C; 548 G; 666 T; 0 U; 1 Other;
Query Match 10.9%; Score 238.4; DB 6; Length 2716;
Best Local Similarity 99.6%; Pred. No. 5.4e-50;
Matches 239; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1730 GGGGAATAGCAGAAATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAGG 1789
Db 1 GGGCAATAGCAGAAATAGGAGCAAGCCAGCAGCTAGTCACTAAGTCACTCAACCAAGG 60
QY 1790 CCTTTTTCCTTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 1849
Db 61 CCTTTTTCCTTGTATCTTTGAGATACATCTATTTCTTAGCGTTTCTGGAGATTACA 120
QY 1850 CATCTCGGTTCCGTTTCTGGGAACCTTCTGATTTATCTCCCGCTCACACAAATAAG 1909
Db 121 CATCTCGGTTCCGTTTCTGGGAACCTTCTGATTTATCTCCCGCTCACACAAATAAG 180
QY 1910 CATTTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 1969
Db 181 CATTTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 240
RESULT 9
AAC00831
ID AAC00831 standard; cDNA; 640 BP.
XX

Db 11183 CAATAAACCCACCCCTATCCAAAAATAAATAAATCAAAATCTTAATAACAATAAATCTCTA 11124
QY 1272 CCAAGATCTCTCTGCTGCACTGCACTCTGTGCACAGGAAGAGGGGAGAGGCAAT 1331
Db 11123 CCAAAATCTCTACTCTCTACCTCTATATAACAATAAATAAATAAATAAATAAATTT 11064
QY 1332 GTTGCTTCTCTTATTTG-----AGCATATTTAAGGGAATATAAGTCTCTATGACTCA 1387
Db 11063 TACCTTTCCCAATTTATTAATAATCCTTATCAACAGCACTACCGATTTCTATAAACCTA 11004
QY 1388 CAGTGAAGTCTGCTGGGTAGGAAGTTCTATCTTGAATCTCTTTCAGTATCGAATTT 1447
Db 11003 TATATAATTAATCTTTATAAAAAAATCTAAATTTAAACTCTCTTCAACAAATATTTAT 10944
QY 1448 TGTAAAGAAAAA 1459
Db 10943 AATCAAAAAA 10932

RESULT 12
ABK28234/c
ID ABK28234 standard; DNA; 15732 BP.

XX AC ABK28234;

XX DT 23-APR-2002 (first entry)

XX DE DNA transcription associated complementary genomic DNA #54.

XX KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.

XX OS Unidentified.

XX PN W0200192565-A2.

XX PD 06-DEC-2001.

XX PF 06-APR-2001; 2001WO-EP003973.

XX PR 06-APR-2000; 2000DE-01019058.

XX PR 07-APR-2000; 2000DE-01019173.

XX PR 30-JUN-2000; 2000DE-01032529.

XX PR 01-SEP-2000; 2000DE-01043826.

XX PA (EPIG-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX DR WPI; 2002-090046/12.

XX PT New nucleic acids or oligomers, useful for diagnosing or treating

XX PT diseases associated with DNA transcription, e.g. immunological disorders,

XX PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or

XX PT cancer.

XX PS Claim 1; SEQ ID NO 108; 32pp; English.

XX CC The invention relates to a nucleic acid, which comprises a segment of the

XX CC chemically pretreated DNA of genes associated with DNA transcription from

XX CC one of 346 sequences, and an oligomer, in particular an oligonucleotide

XX CC or peptide nucleic acid (PNA)-oligomer, that hybridises to or is identical

XX CC to the chemically pretreated DNA of genes associated with DNA

XX CC transcription. The set of oligomer probes are useful for detecting the

XX CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)

CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 15732 BP; 5557 A; 70 C; 2865 G; 7240 T; 0 U; 0 Other;

Query Match 7.5%; Score 165.2; DB 6; Length 15732;
Best Local Similarity 56.8%; Pred. No. 4.3e-31;
Matches 586; Conservative 0; Mismatches 388; Indels 58; Gaps 13;

QY 442 AGTGCCCTTAATATCCCTTCTTCTACCTACTGCTTAAGGCCAACTCAATGTATGCCATATA 501
Db 11919 AATTTCCTTATATCTCTTCTTACCTTTACCTTAAACCTACTCGAAACATCAATATA 11860
QY 502 GAGTACCCCTAAAGCCAGTAAAGATTATCACGAACACCCAGAGATGTTTATGGCCATATT 561
Db 11859 AAAAAAATAA-----ACAAATAACACCCACATATATTAATCAATCAT 11814
QY 562 CCAGATCCCTTGTCTTCTTCTTGTACTACTCTCTTTTGTATAAGTATTTCATATGGTT 621
Db 11813 TCTAATCCCTTAAATTTTCTTAAATTTACCTACTTAAATAATTTAAACA---ATTT 11757
QY 622 TCAATCCAGAAATTCAGCTATTTCACCTTCATGCGCCCTAGTTTGGATTAAACATCAGCTT 681
Db 11756 CCAATCCAACTTTA---ACTCTCTATACTACTAAATTTAAATTAACCAATCACTT 11700
QY 682 TCTTTTGTCTGTAGATGGCTTTGGACAGT-----ACATATTACAGTCTATGGGCCCTT 736
Db 11699 TATTTCTCTATATAAATAACCTTAAATCATTTTCCCTTATTAATTTTAAATCCTTT 11640
QY 737 TCCACTCATGACACATCTGCTTGCAGATCTGCTGCTGTTTAAACACACATCTCTGTC 796
Db 11639 TCCATCATACACCAATTTTACTTTCAATTAATTAATTAATTAATTAATTAATTAATTA 11580
QY 797 TAGTTGATGCCATTTTATTTGACATCTCTGGGTTTACAGCAGCTTAAAGGCGCACCCTGTGTC 856
Db 11579 TAACTAATACCATTTCAATTTAAACAATTTCAAAATTTATAATCTTTTATAATACCCCACTATT 11520
QY 857 CTTAACAAGAGAAAATACACTGCTGAGCAGAGAGATTAAACAGACTCTCTCACACTACGGGA 916
Db 11519 CTTCAAAA-AAAAAAAAAATTACAATACAAAAAATTTATCAAAACCCCTTATCATTTCCAAA 11461
QY 917 AGTTCAGTGCAATTTTCCAGGATTTTCAAGTCTTCGATGCTGTCTGATGTTGAGAGTCCACTA 976
Db 11460 ATTTCAATAC-TATTCAAAATATTTCAAAATTTCAATCAATATAATAACCAA----- 11410
QY 977 TCCACCGTTGTTTATCCAGGAATTAATGCAGAAAATCTGAGCATTCGTGATCTTAATCAAT 1036
Db 11409 -----CTATCCATAAATAATACAAAAAATAAATAAATAAATAAATAAATAAATAA 11361
QY 1037 GAGGTAAAGTAACCTAGTCTATGG---TCAAGAGCAGCAGTGTGTTAAACTCTTTCATAT 1093
Db 11360 AAAATAAAACAACACTCAATTTATATCTCAAAAAAACAACATAATAAATAAATAAATAA 11301
QY 1094 AAATTTAAAGGAAGATTATCATATTAATGACTTTCT--GATCTGGAAGGCCAGGCT 1151
Db 11300 -TATTTAAATAAATAATTTACATTTTCTAATAAATTTTCTACCTCTATCAAACTCAATA 11242
QY 1152 TGTCAAGTTAATAAATTTGATGAATCTTTCAGGTAGTATTTAAACAAATAGCGATTTCA 1211
Db 11241 TATCAATTTAATAACTAAATAAATA-TATAAATACTATTTAAACAA- AACAAATTTCA 11184

QY	1212	TGAAGACATTTCCCCCAACGAGGAGATGCAGTAAAGTATTGGTAGCAATAGATGCTCCTA	1271
Db	11183	CAATAAACCCACCCCTATCCAAAAATAAATCAAAATTCATAACAATAACATCTCTA	11124
QY	1272	CCAAGATCTCCTCCTGCGCACTGTGTCACAGGAAAGGAGGGGGAAGAAGCCAT	1331
Db	11123	CCAAAAATCTCTACTCTCTACCACCTCTATAACAACAAAAAATAAAACATTTTAATATT	11064
QY	1332	GTTCCTCTCTTATATTG-----AGCATCATTTAAGGGAATATAAGTCTCTAATGACTCA	1387
Db	11063	TACCTTTCCCAATTTATATTAATAATCCCTTATTCAACGACTACCGATTTCTAATAACCTA	11004
QY	1388	CATGTAAGTCGTGGTGAAGAAAGTTCTAAATGTTGGAACTCCTTTACGACTATCGAATTT	1447
Db	11003	TATAATAATTAACTTTATAAATAAATCTAATAATTAAACTCCTTCAACAATATTAT	10944
QY	1448	TGTAAGAAAAA	1459
Db	10943	AAATCAAAAAA	10932

RESULT 13
AAS45388
ID AAS45388 standard; DNA: 15732 BP.

XX
AC AAS45388;
XX
XX 18-DEC-2001 (first entry)
DT
XX
XX Chemically pretreated genomic DNA associated with cell cycle #47.
DE

Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging; KW
 human immunodeficiency virus; neurodegenerative disorder; solid tumour; KW
 graft-versus-host disease; glomerular disease; Lewy body disease; cancer; KW
 arthritis; arteriosclerosis; anti-HIV; neuroprotective; antithalamic; KW
 immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds; KW
 PCR primer. KW

XX	Homo sapiens.
OS	
XX	WO200168911-A2.
PN	
XX	20-SEP-2001.
PD	
XX	15-MAR-2001; 2001WO-EF002945.
PF	
XX	15-MAR-2000; 2000DE-01013847.
XX	06-APR-2000; 2000DE-01019058.
PR	07-APR-2000; 2000DE-01019173.
PR	30-JUN-2000; 2000DE-01032529.
PR	01-SEP-2000; 2000DE-01043826.
XX	(EPIG-) EPIGENOMICS AG.
PA	
XX	Olek A, Piepenbrock C, Berlin K;
PI	
XX	WPI: 2001-602751/68.
XX	
DR	

XX	Designing primers and probes for analyzing diseases associated with
PT	cytosine methylation state e.g. arthritis, cancer, aging,
PT	arteriosclerosis comprising fragments of chemically modified genes
PT	associated with cell cycle.
XX	
XX	Claim 1; SEQ ID NO 93; 28pp; English.
XX	
CC	Sequences AAS45296-AAS45250 represent chemically pretreated genomic DNA
CC	molecules associated with the cell cycle and specific PCR primers of the
CC	invention. The sequences are useful for detecting the methylation state
CC	of all CpG dinucleotides in a sequence and therefore for analysing
CC	associated diseases. By analysing cytosine methylations in the pretreated
CC	DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
CC	of existing diseases or the predisposition to specific diseases can be

CC	ascertained. The parameters may be compared to another set of genetic
CC	and/or epigenetic parameters, the differences serving as basis for
CC	diagnosis and/or prognosis events which are disadvantageous to patients.
CC	The sequences of the invention are useful for the diagnosis and therapy
CC	of HIV infection, neurodegenerative disorders, graft-versus-host disease,
CC	aging, glomerular disease, lewy body disease, arthritis,
CC	arteriosclerosis, solid tumours and cancers
XX	
SQ	Sequence 15732 BP; 4638 A; 70 C; 2672 G; 8352 T; 0 U; 0 Other;
	Query Match 7.2%; Score 157.4; DB 4; Length 15732;
	Best Local Similarity 56.9%; Pred. No. 4.1e-29;
	Matches 526; Conservative 0; Mismatches 351; Indels 48; Caps 11;
QY	528 ATCACGAACACCAGAGATGTTATAGCCATATTCAGATCCCTCTTGTCTTTTCTTGAC 587
DB	3886 ATAACTGAATTTATATGTTAGTGGTGTGTTATTTAGATTTTTTGGTTTTTIGAT 3945
QY	588 TTACTCCTTTTGTGATAAAGTTATTCATATGTTTCAAATCCAGAATCAGCTATTCACT 647
DB	3946 TTA-----TTTGTGTTAGTGGGTTAGATAGTTTTTAAATTTAAATTTTAAATTTT 3999
QY	648 TCATGCCCTAGTTTCGATTAACAAATCAGCTTCTTTGCTCTAGATGGCCTTGA 707
DB	4000 TTAATGTTGTTGGTTTCGATTAAGATAATTAGTTTTATTTTTTTTCTAGATGGTTTGA 4059
QY	708 CAGTACATATTAACTGCTAT-----GGGCCCTTCCACTCATGACACATCCTGCTTGC 762
DB	4060 TTAATTTTTTTTATATAGTTGTTTGGGATTTTTTTTAGTATGATATATTTTGTATT 4119
QY	763 AGATCTGCTGGGTTAAACACACATCATCTGCTAGTTGATGCCATTTATTTGACATCC 822
DB	4120 AGATTAGTTAGTTGGTAATTATAGTATGTTGTTGTTAAATGATATTTATTTTGAAT 4179
QY	823 TGGGGTTACAGCAGCTAAAGGCCACCACTGCTCCTTAACAGAGAAATACACTCTGA 882
DB	4180 TAGAGTTATGATATTTTGTAGTGTATTTATGTTTTTATAG-GAAAAAATAATCTAGT 4238
QY	883 GCAGAGGATTTAAACAGACTCCTCACACTACCGAAGTTCAGTGCAATTTTCCAGGATTTTC 942
DB	4239 ATAAAGGTTTATAGTTTTTTATTTATTTTAGAGTTTTAGTGTGTTTAGAATCTTTA 4298
QY	943 AAGTCTTCGATGCTGTCATGTTGAGAGTCCACTATCCACGTTTGTGTTATCCAGGAATAAT 1002
DB	4299 AGGTTTTAGTTATGATGA-----TTAATTTATTTAATAATAAT 4338
QY	1003 GCAGAAAATCTGAGCAATTCGTGAACTCAATCATTTAGGTAAAGTAACCTAGTCACTGCTCA 1062
DB	4339 GTAAAAGTATGAGTAATTTATGAATTTAATAATTAAGGTAAGGTAGTTTAGTTATTATAT 4398
QY	1063 AGAGGACAGCTAGTGTTA--AACTCCTTCATATAAATTTAAAGGAAGATTTACATATTCA 1120
DB	4399 TTAAGAGATTAATTAGTATGAATTTTTTTTAAATTTTGAATAGGATTTATATTTA 4458
QY	1121 TAATGACTTTCTGATCTGG--AAAGCCAGGCTTGTCACTATTATAAATTTGATGAATA 1178
DB	4459 TAATGATTTTTTGTGTTTGTAAAGTTTAGTGTAATTAGTTATTATAAATTTGGTGAATA 4518
QY	1179 CTTTCAGGTAGTATTTAACAATAAGCGATTCATAGAAGACATTTCCCAACACAGAGAT 1238
DB	4519 -TGTAGGTATATTAAATTTAAATAAATAATTTTATAGTAGATTATTTTATTTAAGAGAT 4576
QY	1239 GCAGTAAAGTATTGGTAGCAATAGATGCTCCTACCAAGATCTCCTCCGCACTGCCACT 1298
DB	4577 GAAGTTAAGTTTTGGTAGTAATAGATATTTTATTAAGAT---TTTTGTTTTTGTATT 4633
QY	1299 CTGTGCACAGGAAAGAGGGGAGAGAACCATGTTGCTTCTCTATATTT-----GA 1351
DB	4634 TTGTGAATAGGAAAGAGGGGAGAAATTTTGGTCTTTGTTTTTTTAAATTTATTAAGA 4693
QY	1352 GCATCAATTTAAGGGAATATAAGTCTCTAATGACTCATGTAACTGCTGGGTAAAGAA 1411
DB	4694 TTTTATTTTAAAGCGTTATCGGTTTTTAAATGATTTATATGTTATAGTTAGTTTATTAAGAA 4753

Search completed: March 28, 2004, 01:38:52
Job time : 922 secs

```
RESULT 15
ABX62895
ID   ABX62895 standard; cDNA; 2905 BP.
XX
AC   ABX62895;
XX
DT   25-FEB-2003 (first entry)
XX
DE   Human activated T cell cDNA #11.
XX
KW   T cell; gene; es; differential expression; T cell activation;
KW   antiallergic; cytostatic; immunosuppressive; antimicrobial; gene therapy;
KW   allergy; cancer; graft versus host disease; infection;
KW   autoimmune disorder.
XX
OS   Homo sapiens.
XX
PN   US2002137077-A1.
XX
PD   26-SEP-2002.
XX
PF   25-OCT-2001; 2001US-00002600.
XX
PR   25-OCT-2000; 2000US-0243521P.
XX
PA   (HOPK/) HOPKINS C M.
PA   (PETE/) PETERSON D P.
PA   (COCK/) COCKS B G.
PA   (HAWK/) HAWKINS P R.
XX
PI   Hopkins CM, Peterson DP, Cocks BG, Hawkins PR;
XX
DR   WPI; 2003-102381/09.
XX
PT   New combination comprising several cDNAs that are differentially
PT   expressed in activated T cells, useful for diagnosing, treating, staging
PT   or monitoring treatment for allergy, cancer, infectious and/or autoimmune
PT   disorders.
XX
PS   Claim 1; Page; 180pp; English.
XX
CC   This invention relates to the sequences of several cDNAs that are
CC   differentially expressed in activated T cells. The sequences of the
CC   invention may have antiallergic, cytostatic, immunosuppressive and
CC   antimicrobial activity and may be used in gene therapy. The invention
CC   also comprises a method for screening samples for differentially
CC   expressed genes and a method for detecting these cDNAs by hybridisation.
CC   The methods and compositions of the present invention are useful for
CC   diagnosing, treating, staging or monitoring treatment for allergy,
CC   cancer, chronic graft versus host disease, infectious and/or autoimmune
CC   disorders. The present sequence represents a cDNA of the invention that
CC   is differentially expressed in activated T cells
XX
SQ   Sequence 2905 BP; 1030 A; 593 C; 577 G; 705 T; 0 U; 0 Other;
Query Match      7.2%; Score 157; DB 7; Length 2905;
Best Local Similarity 100.0%; Pred. No. 2.7e-29;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1813 AGATACCTTCATTTCTTAGCGTTTCTCGAGATTACACATCTCGCGTTCGGTTTCTGGG 1872
Db 107 AGATACCTTCATTTCTTAGCGTTTCTCGAGATTACACATCTCGCGTTCGGTTTCTGGG 166
QY 1873 AACCTTACTGATTTATCTCCCCCTCACACAAATAGCATTGATCTCTGATTTCTGAAG 1932
Db 167 AACCTTACTGATTTATCTCCCCCTCACACAAATAGCATTGATCTCTGATTTCTGAAG 226
QY 1933 ATCTCAGAGACTGGACTACTCTTGAAAAATTTCCAG 1969
Db 227 ATCTCAGAGACTGGACTACTCTTGAAAAATTTCCAG 263
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 01:12:30 ; Search time 184 Seconds
(without alignments)
6608.138 Million cell updates/sec

Title: US-09-931-733-1
Perfect score: 2191
Sequence: 1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	76.8	3.5	1359	4	US-09-833-381-997
C 2	74.2	3.4	786431	4	US-09-751-389-3
C 3	73.8	3.4	9377	4	US-09-801-874-3
C 4	71	3.2	99916	4	US-09-816-095-3
C 5	69.2	3.2	694	1	US-08-358-171-25
C 6	69.2	3.2	694	3	US-09-090-947-25
C 7	69.2	3.2	1201	1	US-08-358-171-9
C 8	69.2	3.2	1201	3	US-09-090-947-9
C 9	69.2	3.2	1426	1	US-08-358-171-3
C 10	69.2	3.2	1426	3	US-09-090-947-3
C 11	65.6	3.0	478	4	US-09-495-050A-204
C 12	64.6	2.9	392000	4	US-10-027-983-11
C 13	64	2.9	13187	4	US-09-422-936-61
C 14	63.2	2.9	81001	4	US-09-750-580-1
C 15	61	2.8	84495	4	US-09-797-906-3
C 16	60.4	2.8	72928	3	US-09-009-913-1
C 17	59.8	2.7	38564	4	US-09-734-673-3
C 18	59.4	2.7	44848	4	US-09-435-739-42
C 19	58	2.6	168575	4	US-09-426-290-1
C 20	56.8	2.6	43795	3	US-08-742-185-101
C 21	55.6	2.5	1849	4	US-09-023-655-1232
C 22	55.6	2.5	246240	2	US-08-724-394A-20
C 23	55.6	2.5	246240	2	US-08-724-394A-21
C 24	55.6	2.5	246240	2	US-08-724-394A-22
C 25	54.6	2.5	2012	1	US-07-866-560-3
C 26	54.6	2.5	2012	1	US-08-077-673-3
C 27	54.6	2.5	2012	1	US-08-478-992-3

28	54.6	2.5	2012	3	US-09-105-298-3	Sequence 3, Appli
29	54.6	2.5	2012	3	US-08-706-281A-7	Sequence 7, Appli
30	54.6	2.5	2012	3	US-09-097-231-7	Sequence 7, Appli
31	54.6	2.5	2012	4	US-09-542-122-3	Sequence 7, Appli
32	54.6	2.5	2012	4	US-09-353-099-7	Sequence 7, Appli
33	53.8	2.5	319608	4	US-09-539-333D-1	Sequence 1, Appli
34	53.8	2.5	319608	4	US-09-679-409-1	Sequence 1, Appli
35	52.8	2.4	116592	4	US-09-818-512-3	Sequence 3, Appli
36	51.4	2.3	1001	4	US-09-671-317-355	Sequence 355, App
37	51.4	2.3	1001	4	US-09-671-317-357	Sequence 357, App
C 38	51	2.3	90541	4	US-09-759-359A-3	Sequence 3, Appli
C 39	49.2	2.2	695	4	US-09-200-934-6	Sequence 6, Appli
C 40	49	2.2	87350	3	US-08-781-891-79	Sequence 79, Appl
C 41	49	2.2	87350	4	US-09-618-166-79	Sequence 79, Appl
C 42	49	2.2	87543	4	US-09-791-211-3	Sequence 3, Appli
C 43	48.4	2.2	55827	4	US-09-813-133A-3	Sequence 3, Appli
C 44	48	2.2	4011	1	US-08-121-057-3	Sequence 3, Appli
C 45	48	2.2	4011	2	US-08-509-187D-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-833-381-997/c
; Sequence 997, Application US/09833381
; Patent No. 6672186
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: No. 6672186el Nucleic Acid and Protein Homologs
; FILE REFERENCE: 5800-119
; CURRENT APPLICATION NUMBER: US/09/833,381
; CURRENT FILING DATE: 2001-04-11
; PRIOR APPLICATION NUMBER: 09/516,448
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 2050
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 997
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-833-381-997

Query Match Similarity	3.5%	Score	76.8	DB	4	Length	1359
Best Local Similarity	72.4%	Pred. No.	1.2e-11				
Matches	113	Conservative	0	Mismatches	42	Indels	1
Gaps	1						
QY	1460	ACTTATTGTTGTTTATATACATTTATGAGATACAGTGTAAATTCGTTACATGGATATAT	1519				
Db	1315	ATTTATTATTGTAATACATTTAAAGGGTACAGCATAATTTTGTTCATGGATATAT	1256				
QY	1520	GCCATATTGGTGAAGTCAGAGATTCAGTGTGCACATCACCGGAAATGTTTAACTGTAC	1579				
Db	1255	CACAAAGTGGAGAGCTGTGAGCTTTAGTTTATCCATCACCTAAATATTGTACATTATAC	1196				
QY	1580	CCATTAAAGTAAATTCCTCATCCCATTTCCCTCA	1615				
Db	1195	CCAGT-AGGTAATTTCTCATCCCACTCCCA	1161				

RESULT 2

US-09-751-389-3/c
; Sequence 3, Application US/09751389
; Patent No. 6630334
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001067
; CURRENT APPLICATION NUMBER: US/09/751,389
; CURRENT FILING DATE: 2001-01-02
; NUMBER OF SEQ ID NOS: 8

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 786431
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1) ... (786431)
; OTHER INFORMATION: n = A,T,C or G
US-09-751-389-3

Query Match 3.4%; Score 74.2; DB 4; Length 786431;
Best Local Similarity 65.0%; Pred. No. 2.1e-09;
Matches 141; Conservative 0; Mismatches 73; Indels 3; Gaps 2;

QY 1446 TTGTGTAAGAAAAAAGCTTATTTGTGTTTATATACATTATATGAGATACAAAGTGTAAATTCG 1505
DB 170208 TTTTAAATGATAAATCTTTTGTTACTTGTATATAATGTATAGATAACAAGTATAATTTA 170149

QY 1506 TTACATGATATATGCCATATGTGTGAAGTCAGAGATTTTCAGTGTGCACATCACCGGAA 1565
DB 170148 TTACACTAATATATGTCATATGTGTGAAGTCAGGGCTTT-AGTGTACTCATCACCGGA 170090

QY 1566 AATGTAACTGTACCCATTAAAGTAATTTCTCATCCCCATTTCCCTCCACCATGTCTC 1625
DB 170089 AATGTACATTGCACC--ATTAAAGTAATTTCTCATCCACCCCTTTTACCCTTTCAC 170032

QY 1626 CAGCCCTTGCAGGAAAGCTGTTCAATTTTCTCTCACTA 1662
DB 170031 CTTTCCAAGTCTCAATGTTCTATCATTTCCACAGTGTA 169995

RESULT 3
US-09-801-874-3
; Sequence 3, Application US/09801874
; Patent No. 6582935
; GENERAL INFORMATION:
; APPLICANT: VAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN AMINOTRANSFERASE
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: AMINOTRANSFERASE PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL000615
; CURRENT APPLICATION NUMBER: US/09/801.874
; CURRENT FILING DATE: 8001-03-09
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9377
; TYPE: DNA
; ORGANISM: Human
US-09-801-874-3

Query Match 3.4%; Score 73.8; DB 4; Length 9377;
Best Local Similarity 70.2%; Pred. No. 2.4e-10;
Matches 113; Conservative 0; Mismatches 47; Indels 1; Gaps 1;

QY 1462 TTATTTGTGTTTATATACATTATGAGATACAAAGTGTAAATTCGTTACATGATATATCG 1521
DB 1570 TTTTAAATTTGTATATAATGTGTGAAGTACAAAGTGTAAATTTGTTAGATGCATACATCG 1629

QY 1522 CATATGTGTGAAGTCAGAGATTTTCAGTGTGCACATCACCCGAAAAAGTTAACTGTACCC 1581
DB 1630 CATAGTGTGTGAAGTCAGGGCTTTTAGGGTATCCCAATACCCCAATGTACATTGTATCT 1689

QY 1582 ATTAAAGTAATTTCTCATCCCCATTTCCCTCCACCATG 1622
DB 1690 GTT-AGTAATCTCCCATATACCTTGCCTCTCATCAAG 1729

RESULT 4
US-09-816-095-3
; Sequence 3, Application US/09816095

```

REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
OTHER INFORMATION: /note= Exon 7 corresponds to
OTHER INFORMATION: nucleotides 39...670
US-08-358-171-25

Query Match 3.2%; Score 69.2; DB 1; Length 694;
Best Local Similarity 72.5%; Pred. No. 1.2e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1476 ATACATTTATGAGATACAAGTGTAAATTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
DB 423 ATAAATGTAAGGGGTACAGTGCAGTTTGTTCATGATAGATTGCCTAGTGGTGAAGT 482

QY 1536 CAGAGATTTTCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCCATTAAAGTAAATTC 1595
DB 483 CTGGGCTTTTAGTGTAAACATCACCTTAATATATACGTTGTACCCATT-AAAGTTATTC 541

QY 1596 TCATCCCCCATTTCCCTCACC 1617
DB 542 TCATCCCTCACCCCTCCACC 563

RESULT 6
US-09-050-947-25
Sequence 25, Application US/09090947
Patent No. 6008338
GENERAL INFORMATION:
APPLICANT: FONG, Henry K.W.
TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/090,947
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/358,171
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 39...797
US-08-358-171-9

LENGTH: 694 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
OTHER INFORMATION: /note= Exon 7 corresponds to
OTHER INFORMATION: nucleotides 39...670
US-09-090-947-25

Query Match 3.2%; Score 69.2; DB 3; Length 694;
Best Local Similarity 72.5%; Pred. No. 1.2e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1476 ATACATTTATGAGATACAAGTGTAAATTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
DB 423 ATAAATGTAAGGGGTACAGTGCAGTTTGTTCATGATAGATTGCCTAGTGGTGAAGT 482

QY 1536 CAGAGATTTTCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCCATTAAAGTAAATTC 1595
DB 483 CTGGGCTTTTAGTGTAAACATCACCTTAATATATACGTTGTACCCATT-AAAGTTATTC 541

QY 1596 TCATCCCCCATTTCCCTCACC 1617
DB 542 TCATCCCTCACCCCTCCACC 563

RESULT 7
US-08-358-171-9
Sequence 9, Application US/08358171
Patent No. 5763578
GENERAL INFORMATION:
APPLICANT: FONG, Henry K.W.
TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,171
FILING DATE: 16-DEC-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 39...797
US-08-358-171-9

Query Match 3.2%; Score 69.2; DB 1; Length 1201;

Best Local Similarity 72.5%; Pred. No. 1.6e-09; Mismatches 38; Indels 1; Gaps 1;
Matches 103; Conservative 0;

QY 1476 ATACATTATGAGATACAAAGTGTAAATTCGTTACATGGATATATGCCATATTTGGTGAAGT 1535
Db 1053 ATAAATGTAAGGGGTACAGTGCAGTTTGTACATGGATAGATTGCCCTAGTGGTGAAGT 1112

QY 1536 CAGAGATTTCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCTATGCCATTAAGTAATTC 1595
Db 1113 CTGGCTTTTAGTGTAAACCATCACCTAATAATATACGTTTACCCATT-AAGTTATTTC 1171

QY 1596 TCATCCCCCAATTCCTCCACC 1617
Db 1172 TCATCCCTCACCCCTCCACC 1193

RESULT 8
US-09-090-947-9
; Sequence 9, Application US/09090947
; Patent No. 6008338
; GENERAL INFORMATION:
; APPLICANT: FONG, Henry K.W.
; TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/090,947
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/358,171
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: FONG=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 39..797
US-09-090-947-9

Query Match 3.2%; Score 69.2; DB 3; Length 1201;
Best Local Similarity 72.5%; Pred. No. 1.6e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;

QY 1476 ATACATTATGAGATACAAAGTGTAAATTCGTTACATGGATATATGCCATATTTGGTGAAGT 1535
Db 1053 ATAAATGTAAGGGGTACAGTGCAGTTTGTACATGGATAGATTGCCCTAGTGGTGAAGT 1112

QY 1536 CAGAGATTTCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCTATGCCATTAAGTAATTC 1595

Db 1113 CTGGCTTTTAGTGTAAACCATCACCTAATAATATACGTTTACCCATT-AAGTTATTTC 1171

QY 1596 TCATCCCCCAATTCCTCCACC 1617
Db 1172 TCATCCCTCACCCCTCCACC 1193

RESULT 9
US-08-358-171-3
; Sequence 3, Application US/08358171
; Patent No. 5763578
; GENERAL INFORMATION:
; APPLICANT: FONG, Henry K.W.
; TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/358,171
; FILING DATE: 16-DEC-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: YUN, Allen C.
; REGISTRATION NUMBER: 37,971
; REFERENCE/DOCKET NUMBER: FONG=2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1426 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 39..911
US-08-358-171-3

Query Match 3.2%; Score 69.2; DB 1; Length 1426;
Best Local Similarity 72.5%; Pred. No. 1.7e-09;
Matches 103; Conservative 0; Mismatches 36; Indels 1; Gaps 1;

QY 1476 ATACATTATGAGATACAAAGTGTAAATTCGTTACATGGATATATGCCATATTTGGTGAAGT 1535
Db 1167 ATAAATGTAAGGGGTACAGTGCAGTTTGTGTACATGGATAGATTGCCCTAGTGGTGAAGT 1226

QY 1536 CAGAGATTTCAGTGTGCACATCACCCGAAAAATGTTAACTGTACCTATGCCATTAAGTAATTC 1595
Db 1227 CTGGCTTTTAGTGTAAACCATCACCTAATAATATACGTTTACCCATT-AAGTTATTTC 1285

QY 1596 TCATCCCCCAATTCCTCCACC 1617
Db 1286 TCATCCCTCACCCCTCCACC 1307

RESULT 10
US-09-090-947-3
; Sequence 3, Application US/09090947

Patent No. 6008338
GENERAL INFORMATION:
APPLICANT: FONG, Henry K.W.
TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/09/090,947
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA: 08/358,171
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: YUN, Allen C.
REGISTRATION NUMBER: 37,971
REFERENCE/DOCKET NUMBER: FONG-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 39...911
US-09-090-947-3
Query Match 3.2%; Score 69.2; DB 3; Length 1426;
Best Local Similarity 72.5%; Pred. No. 1.7e-09;
Matches 103; Conservative 0; Mismatches 38; Indels 1; Gaps 1;
Qy 1476 ATACATTATGAGATACAGTGTAAATTCGTTACATGATATATGCCATATTTGGTGAAGT 1535
Db 1167 ATAATGTAAAGGGGTACAGTGCAGTGTGTTTGTACATGATAGATTGCTAGTGTGAAGT 1226
Qy 1536 CAGAGATTTCAGTGTGCATACACCCGAAATATGTTAACTGTACCCATTAAGTAATTC 1595
Db 1227 CTGGGCTTTAGTGTAAACATCACCCCTAATATATACCTGTACCCATTAGTTATTC 1285
Qy 1596 TCATCCCCCATTCGCCCTCACC 1617
Db 1286 TCATCCCTCACCCCTCCACC 1307
RESULT 11
US-09-495-050A-204
Sequence 204, Application US/09495050A
Patent No. 6492505
GENERAL INFORMATION:
APPLICANT: Roopa, Reddy
APPLICANT: Guegler, Karl, J.
APPLICANT: Au-Young, Janice
TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING MEMBRANE-ASSOCIATED
FILE REFERENCE: PA-0013 US
CURRENT APPLICATION NUMBER: US/09/495,050A

CURRENT FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/118,318
PRIOR FILING DATE: February 1, 1999
NUMBER OF SEQ ID NOS: 305
SOFTWARE: PERL Program
SEQ ID NO 204
LENGTH: 478
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6492505 2375491CT1
US-09-495-050A-204
Query Match 3.0%; Score 65.6; DB 4; Length 478;
Best Local Similarity 51.9%; Pred. No. 9.9e-09;
Matches 104; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Qy 1450 TAAGAAAAAACTTAATTTGTTTATATATACATTTATGAGATACAAAGTGTAAATTCGTTAC 1509
Db 263 TAAATTTTATTTTACTTTTTCAGTATAAATTTAAGGGGAACAAGTGCAGTTTTCGTTAC 322
Qy 1510 ATGGATATATGCCATATATTTGTTGTAAGTCAGAGATTTGAGTGTGCATCACCAGGAAATG 1569
Db 323 ATGGATATATTTGCTAGTGTGAAGCTGGGCCCTTTAATGGGACCATCGCCCGGATAGTG 382
Qy 1570 TTAACGTACCCATTAAAGTAATTTCTCATCCCCCATTTCCCTCACC 1617
Db 383 TGCATTGTACCCCATATGAATTTACTCAACCCCTACTCACCCTACCACC 430
RESULT 12
US-10-027-983-11
Sequence 11, Application US/10027983
Patent No. 6617162
GENERAL INFORMATION:
APPLICANT: Kenneth W. Dobie
APPLICANT: Mark P. Roach
TITLE OF INVENTION: ANTISENSE MODULATION OF ESTROGEN RECEPTOR ALPHA EXPRESSION
FILE REFERENCE: RFS-0340
CURRENT APPLICATION NUMBER: US/10/027,983
CURRENT FILING DATE: 2001-12-18
NUMBER OF SEQ ID NOS: 98
SEQ ID NO 11
LENGTH: 392000
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 137740
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 137742
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (138122)...(138221)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 145507
OTHER INFORMATION: unknown
NAME/KEY: unsure
LOCATION: 151967
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (151967)...(1542066)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: unsure
LOCATION: 154217
OTHER INFORMATION: unknown
NAME/KEY: misc_feature
LOCATION: (164037)...(164136)
OTHER INFORMATION: n = A,T,C or G
NAME/KEY: misc_feature

```

; LOCATION: (174657)...(174756)
; NAME/KEY: n = A,T,C or G
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (186224)...(186323)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (195242)...(195341)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: unsure
; LOCATION: 202703
; OTHER INFORMATION: unknown
; NAME/KEY: misc feature
; LOCATION: (202771)...(202870)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (206246)...(215602)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; LOCATION: (218126)...(218225)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (220360)...(220459)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; LOCATION: (222717)...(222816)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (223981)...(224080)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; LOCATION: (227487)...(227586)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (230157)...(230256)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; LOCATION: (232299)...(232398)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; NAME/KEY: misc feature
; LOCATION: (236552)...(236651)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; LOCATION: (238789)...(248788)
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: n = A,T,C or G
; LOCATION: (118288)...(119101)
; OTHER INFORMATION: exon 1C
; NAME/KEY: exon:intron junction
; LOCATION: (151129)...(151130)
; OTHER INFORMATION: exon 5:intron 5
; NAME/KEY: exon:intron junction
; LOCATION: (299248)...(299249)
; OTHER INFORMATION: exon 9:intron 9
; NAME/KEY: exon:intron junction
; LOCATION: (348578)...(348579)
; OTHER INFORMATION: exon 10:intron 10
; NAME/KEY: intron
; LOCATION: (348579)...(381838)
; OTHER INFORMATION: intron 10
; NAME/KEY: intron:exon junction
; LOCATION: (386185)...(386186)
; OTHER INFORMATION: intron 11:exon 12
; US-10-027-983-11

```

```

Query Match 2.9%; Score 64.6; DB 4; Length 392000;
Best Local Similarity 59.6%; Pred. No. 7.4e-07;
Matches 109; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 1446 TTGTGAAGAAAAAAGTATTTTGTGTTTATATACATTTATGATACATACAGTGTAAATTCG 1505
DB 258633 TTTTACTTATTTATTTATTTATCCACAGGTTTTTGAAGAAACAGGTGATGTTGG 258692
QY 1506 TTACATGATATATGCCATATTGTTGAAGTCAGAGATTTTCAGTGTGCATCATCCCGAAA 1565

```

```

Db 258693 TTACATGAATAAGTCTTTTAGTGTGATTCTGAGATTTTGGTGATCATCCACGAGC 258752
QY 1566 AATGTTAACTGTACCCATTAAGTAATTTCTCATCCCCCATTTCCCTCACCACATGCTC 1625
Db 258753 GGTGTACACTGTATCTAGTTTGTAAATTTTTTATCCCTCACTCCTTCCCTTCTTC 258812
QY 1626 CAG 1628
Db 258813 CCG 258815

RESULT 13
US-09-422-936-61
; Sequence 61, Application US/09422936
; Patent No. 6465213
; GENERAL INFORMATION:
; APPLICANT: Ekstrand, Jonas
; TITLE OF INVENTION: NEW NUCLEOTIDE SEQUENCES
; FILE REFERENCE: 06275-165002
; CURRENT APPLICATION NUMBER: US/09/422,936
; CURRENT FILING DATE: 1999-10-22
; PRIOR APPLICATION NUMBER: US 09/242,608
; PRIOR FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: PCT/SE98/01947
; PRIOR FILING DATE: 1998-10-27
; PRIOR APPLICATION NUMBER: SWEDEN 9703914-2
; PRIOR FILING DATE: 1997-10-27
; PRIOR APPLICATION NUMBER: SWEDEN 9800864-2
; PRIOR FILING DATE: 1998-03-16
; PRIOR APPLICATION NUMBER: SWEDEN 9802575-2
; PRIOR FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 13187
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(13187)
; OTHER INFORMATION: n = A, T, G, or C
; NAME/KEY: misc feature
; LOCATION: (1)...(15652)
; OTHER INFORMATION: r = G or A; y = T/U or C; m = A or C;
; OTHER INFORMATION: k = G or T/U; s = G or C; w = A or T/U;
; OTHER INFORMATION: b = G, C, or T/U; d = A, G, or T/U;
; OTHER INFORMATION: h = A, C, or T/U; v = A, G, or C
; NAME/KEY: intron
; LOCATION: (1)...(1924)
; NAME/KEY: exon
; LOCATION: (1925)...(2167)
; NAME/KEY: intron
; LOCATION: (2168)...(2795)
; NAME/KEY: exon
; LOCATION: (2796)...(2859)
; NAME/KEY: intron
; LOCATION: (2860)...(4408)
; NAME/KEY: exon
; LOCATION: (4409)...(4486)
; NAME/KEY: intron
; LOCATION: (4487)...(6030)
; NAME/KEY: exon
; LOCATION: (6031)...(6181)
; NAME/KEY: intron
; LOCATION: (6182)...(6678)
; NAME/KEY: exon
; LOCATION: (6679)...(6811)
; NAME/KEY: intron
; LOCATION: (6812)...(8194)
; NAME/KEY: exon
; LOCATION: (8195)...(8311)
; NAME/KEY: intron

```

LOCATION: (8312)...(8408)
NAME/KEY: exon
LOCATION: (8409)...(8516)
NAME/KEY: intron
LOCATION: (8517)...(8928)
NAME/KEY: exon
LOCATION: (8929)...(9022)
NAME/KEY: intron
LOCATION: (9023)...(9716)
NAME/KEY: exon
LOCATION: (9717)...(9844)
NAME/KEY: intron
LOCATION: (9845)...(10424)
NAME/KEY: exon
LOCATION: (10425)...(10553)
NAME/KEY: intron
LOCATION: (10554)...(10775)
NAME/KEY: exon
LOCATION: (10776)...(10919)
NAME/KEY: intron
LOCATION: (10920)...(11700)
NAME/KEY: exon
LOCATION: (11701)...(13187)
US-09-422-936-61

Query Match 2.98; Score 64; DB 4; Length 13187;
Best local Similarity 68.8%; Pred. No. 1.7e-07;
Matches 88; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy 1458 AAATCTATTGTTGTTATATACATTTATGAGATACAAAGTGTAAATTCGTTACATGGATAT 1517
Db 4960 AAAAATTTCTATTTTATGTAATTTAGAGGTAGAGTGCAGTTGTTACATGAATAT 5019

Qy 1518 ATGCCATATTGGTGAAGTCAGAGATTTTCAGTGTGCACATCACCCGAAAAATGTTAACTGT 1577
Db 5020 ATTGCATAGTGTGTAAGTCCGGGGTTTGTAGTGCCTGTACCCGACAGTGCACCTTGT 5079

Qy 1578 ACCCATTA 1585
Db 5080 ACCTAATA 5087

RESULT 14

US-09-750-580-1
Sequence 1, Application US/09750580
Patent No. 6455280
GENERAL INFORMATION:
APPLICANT: Yen, Frances
APPLICANT: Denison, Blake
APPLICANT: Bour, Barbara
APPLICANT: Bihain, Bernard
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Duclert, Aymeric
APPLICANT: Bougueleret, Lydie
APPLICANT: Ebbets-Reed, Dana
APPLICANT: Salter-Cid, Luisa
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITING NEOPLASTIC CELL GROWTH
FILE REFERENCE: 89.US2.CIP
CURRENT APPLICATION NUMBER: US/09/750,580
CURRENT FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 09/599,362
PRIOR FILING DATE: 2000-06-21
PRIOR APPLICATION NUMBER: PCT/TB00/0101
PRIOR FILING DATE: 2000-06-21
PRIOR APPLICATION NUMBER: PCT/TB99/02058
PRIOR FILING DATE: 1999-12-20
PRIOR APPLICATION NUMBER: US 49/469/099
PRIOR FILING DATE: 1999-12-21
PRIOR APPLICATION NUMBER: US 60/113,686
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: US 60/141,032
PRIOR FILING DATE: 1999-06-25
NUMBER OF SEQ ID NOS: 6

SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 81001
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 10946..12946
OTHER INFORMATION: 5'regulatory region
NAME/KEY: exon
LOCATION: 12947..12958
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 13470..13526
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 13641..13752
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 14271..15968
OTHER INFORMATION: exon 4
NAME/KEY: misc feature
LOCATION: 15969..17969
OTHER INFORMATION: 3'regulatory region
NAME/KEY: allele
LOCATION: 1239
OTHER INFORMATION: 20-828-311 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 12347
OTHER INFORMATION: 17-42-319 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 15241
OTHER INFORMATION: 17-41-250 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 42218
OTHER INFORMATION: 20-841-149 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 45442
OTHER INFORMATION: 20-842-115 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 77058
OTHER INFORMATION: 20-853-415 : polymorphic base C or T
NAME/KEY: primer bind
LOCATION: 929..949
OTHER INFORMATION: 20-828.pu
NAME/KEY: primer bind
LOCATION: 1357..1377
OTHER INFORMATION: 20-828.rp complement
NAME/KEY: primer bind
LOCATION: 12029..12050
OTHER INFORMATION: 17-42.pu
NAME/KEY: primer bind
LOCATION: 12581..12603
OTHER INFORMATION: 17-42.rp complement
NAME/KEY: primer bind
LOCATION: 14992..15012
OTHER INFORMATION: 17-41.pu
NAME/KEY: primer bind
LOCATION: 15460..15482
OTHER INFORMATION: 17-41.rp complement
NAME/KEY: primer bind
LOCATION: 42070..42090
OTHER INFORMATION: 20-841.pu
NAME/KEY: primer bind
LOCATION: 42572..42591
OTHER INFORMATION: 20-841.rp complement
NAME/KEY: primer bind
LOCATION: 45328..45347
OTHER INFORMATION: 20-842.pu
NAME/KEY: primer bind
LOCATION: 45863..45883
OTHER INFORMATION: 20-842.rp complement
NAME/KEY: primer bind

Mon Mar 29 09:55:00 2004

LOCATION: 76644..76664
; OTHER INFORMATION: 20-853.pu
; NAME/KEY: primer bind
LOCATION: 77166..77185
; OTHER INFORMATION: 20-853.rp complement
; NAME/KEY: primer bind
LOCATION: 1220..1238
; OTHER INFORMATION: 20-828-311.mis
; NAME/KEY: primer bind
LOCATION: 1240..1258
; OTHER INFORMATION: 20-828-311.mis complement
; NAME/KEY: primer bind
LOCATION: 12328..12346
; OTHER INFORMATION: 17-42-319.mis
; NAME/KEY: primer bind
LOCATION: 12348..12366
; OTHER INFORMATION: 17-42-319.mis complement
; NAME/KEY: primer bind
LOCATION: 15222..15240
; OTHER INFORMATION: 17-41-250.mis
; NAME/KEY: primer bind
LOCATION: 15242..15260
; OTHER INFORMATION: 17-41-250.mis complement
; NAME/KEY: primer bind
LOCATION: 42199..42217
; OTHER INFORMATION: 20-841-149.mis
; NAME/KEY: primer bind
LOCATION: 42219..42237
; OTHER INFORMATION: 20-841-149.mis complement
; NAME/KEY: primer bind
LOCATION: 45423..45441
; OTHER INFORMATION: 20-842-115.mis
; NAME/KEY: primer bind
LOCATION: 45443..45461
; OTHER INFORMATION: 20-842-115.mis complement
; NAME/KEY: primer bind
LOCATION: 77059..77077
; OTHER INFORMATION: 20-853-415.mis
; NAME/KEY: primer bind
LOCATION: 1227..1251
; OTHER INFORMATION: 20-828-311.probe
; NAME/KEY: misc binding
LOCATION: 12335..12359
; OTHER INFORMATION: 17-42-319.probe
; NAME/KEY: misc binding
LOCATION: 15229..15253
; OTHER INFORMATION: 17-41-250.probe
; NAME/KEY: misc binding
LOCATION: 42206..42230
; OTHER INFORMATION: 20-841-149.probe
; NAME/KEY: misc binding
LOCATION: 45430..45454
; OTHER INFORMATION: 20-842-115.probe
; NAME/KEY: misc binding
LOCATION: 77046..77070
; OTHER INFORMATION: 20-853-415.probe
US-09-750-580-1

Query Match 2.9%; Score 63.2; DB 4; Length 81001;
Best Local Similarity 61.5%; Pred. No. 7.8e-07;
Matches 118; Conservative 0; Mismatches 73; Indels 1; Gaps 1;

QY 1445 TTTTGTAAAGAAAAAAGCTTTTGTGTTTATATACATTTATGAGATACAAGTGTAAATTC 1504
Db 67490 TTATTTTGAATATATATCTTATTTGTATAAAATGATGAATAGAGTGCATTTT 67549

QY 1505 GTTACATGGATATATGCGATTTGGTGAAGTCAGAGATTTTCAGTGTGCATCACCCGAA 1564
Db 67550 GTTCCATGCAGCTTGTGTAGTGAATTCAGGGCTTTTAGGATATCCATCACTTGA 67609

QY 1565 AAATGTTAACTGTAGCCATTAAAGTAATTTCTCATCCCCCAATTTGCCCTCACCACATGCT 1624
Db 67610 TAATGTACATAAGGCCCA-TGAAGTAATTTCTCATCATTCACCCCCCTCCACCTCTCA 67668

QY 1625 CCAGCCCTTGCC 1636
Db 67669 CCCTTCCAAGTC 67680

RESULT 15
US-09-797-906-3
; Sequence 3, Application US/09797906
; Patent No. 6329188
; GENERAL INFORMATION:
; APPLICANT: Zianghe YAN, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL001151CIP
; CURRENT APPLICATION NUMBER: US/09/797,906
; CURRENT FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 84495
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(84495)
; OTHER INFORMATION: n = A,T,C or G
US-09-797-906-3

Query Match 2.8%; Score 61; DB 4; Length 84495;
Best Local Similarity 64.5%; Pred. No. 3.3e-06;
Matches 91; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1445 TTTTGTAAAGAAAAAAGCTTTTGTGTTTATATACATTTATGAGATACAAGTGTAAATTC 1504
Db 7187 TTTCTTATTTATTTATTTATTTTCAATAGTTTTTGGGGGTACAGGTAGTTTTC 7246

QY 1505 GTTACATGGATATATGCGATTTGGTGAAGTCAGAGATTTTCAGTGTGCATCACCCGAA 1564
Db 7247 GTTACATGGATATATGCGATTTTGAAGTGAATTTAGTGCATCACCCCAAG 7306

QY 1565 AAATGTTAACTGTAGCCATTA 1585
Db 7307 CAGTGTACATTTGTACCCAGTA 7327

Search completed: March 28, 2004, 05:41:05
Job time : 193 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 03:14:35 ; Search time 791 Seconds
(without alignments)
10312.407 Million cell updates/sec

Title: US-09-931-733-1
 Perfect score: 2191
 Sequence: 1 tacataattgagagatttta.....tgcgactgaacaaagaaaa 2191

Scoring table: IDENTITY_NUC
Gapop 10.0 . Gapext 1.0

Searched: 2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters: 4917892

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:★

```

1: /cgn2_6/pdata1/2/pubnpna/us07_PUBCOMB.seq.*
2: /cgn2_6/pdata1/2/pubnpna/PCT_NEW PUB.seq.*
3: /cgn2_6/pdata1/2/pubnpna/us06_NEW PUB.seq.*
4: /cgn2_6/pdata1/2/pubnpna/us06_PUBCOMB.seq.*
5: /cgn2_6/pdata1/2/pubnpna/us07_NEW PUB.seq.*
6: /cgn2_6/pdata1/2/pubnpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/pdata1/2/pubnpna/us08_NEW PUB.seq.*
8: /cgn2_6/pdata1/2/pubnpna/us09_PUBCOMB.seq.*
9: /cgn2_6/pdata1/2/pubnpna/us09A_PUBCOMB.seq.*
10: /cgn2_6/pdata1/2/pubnpna/us09B_PUBCOMB.seq.*
11: /cgn2_6/pdata1/2/pubnpna/us09C_PUBCOMB.seq.*
12: /cgn2_6/pdata1/2/pubnpna/us09_NEW PUB.seq.*
13: /cgn2_6/pdata1/2/pubnpna/us10A_PUBCOMB.seq.*
14: /cgn2_6/pdata1/2/pubnpna/us10B_PUBCOMB.seq.*
15: /cgn2_6/pdata1/2/pubnpna/us10C_PUBCOMB.seq.*
16: /cgn2_6/pdata1/2/pubnpna/us10_NEW PUB.seq.*
17: /cgn2_6/pdata1/2/pubnpna/us60_NEW PUB.seq.*
18: /cgn2_6/pdata1/2/pubnpna/us60_PUBCOMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARY

Result No.	Score	Query Match	Length	DB	ID	Description
1	240	11.0	2709	15	US-10-133-937-84	Sequence 84, Appl
2	240	11.0	2709	15	US-10-373-801-17	Sequence 17, Appl
3	240	11.0	2709	15	US-10-159-563-84	Sequence 84, Appl
4	238.4	10.9	1254	14	US-10-084-817-69	Sequence 69, Appl
5	238.4	10.9	2716	14	US-10-240-965-259	Sequence 259, App
6	228.8	10.4	2736	15	US-10-264-049-498	Sequence 498, App
C 7	165.2	7.5	15732	14	US-10-239-676-96	Sequence 96, Appl
C 8	165.2	7.5	15732	14	US-10-240-453-108	Sequence 108, Appl
9	157.4	7.2	15732	14	US-10-239-676-95	Sequence 95, Appl
10	157.4	7.2	15732	14	US-10-240-453-107	Sequence 107, App
11	157	7.2	2070	15	US-10-108-260-450	Sequence 450, App
12	157	7.2	2905	13	US-10-002-600-11	Sequence 11, Appl
13	157	7.2	2905	14	US-10-084-817-70	Sequence 70, Appl
14	157	7.2	2906	10	US-09-971-393-88	Sequence 88, Appl
15	106.4	4.9	1892	12	US-10-433-544-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

```

RES001 1
US-10-133-937-84
/ Sequence 84, Application US/1013937
/ Publication NO. US2003020728A1
/ GENERAL INFORMATION:
/ APPLICANT: Khan, Javed
/ APPLICANT: Ringner, Markus
/ APPLICANT: Peterson, Carsten
/ APPLICANT: Melczar, Paul
/ TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING,
/ TITLE OF INVENTION: DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND
/ TITLE OF INVENTION: OTHER BIOLOGICAL STATES
/ FILE REFERENCE: 11613.56US01
/ CURRENT APPLICATION NUMBER: US/10/133,937
/ CURRENT FILING DATE: 2002-11-04
/ NUMBER OF SEQ ID NOS: 99
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 84
/ LENGTH: 2709
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-133-937-84

```

	Query Match	11.0%;	Score 240;	DB 15;	Length 2709;
	Best Local Similarity	100.0%;	Pred. No. 7.3e-51;		
	Matches 240;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1730	GGGAATAGCAGAAATAGGACGACGCGACTAGTCAGCTAACTAAGTGACTCTAACCAAGG	1789		
Db	1	GGGAATAGCAGAAATAGGACGACGCGACTAGTCAGCTAACTAAGTGACTCTAACCAAGG	60		
QY	1790	CCCTTTTTCCTTGTATCTTTTGCAGATACTTCATTTTCTTAGCGTTTCTGGAGATTACAA	1849		
Db	61	CCCTTTTTCCTTGTATCTTTTGCAGATACTTCATTTTCTTAGCGTTTCTGGAGATTACAA	120		
QY	1850	CAATCTGCGGTTCCGFTTCTGGGAACTTTACTGATTTATCTCCCCCTCACACAAATAAG	1909		
Db	121	CAATCTGCGGTTCCGFTTCTGGGAACTTTACTGATTTATCTCCCCCTCACACAAATAAG	180		


```
; OTHER INFORMATION: Template ID: 040652.36
US-10-002-600-11

Query Match
Best Local Similarity 7.2%; Score 157; DB 13; Length 2905;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1813 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 1872
DB 107 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 166

QY 1873 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTGATTCCTGCAATTCCTGAAG 1932
DB 167 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTGATTCCTGCAATTCCTGAAG 226

QY 1933 ATCTCAAGATCTGGACTACTGTGTAATAAATTTCCAG 1969
DB 227 ATCTCAAGATCTGGACTACTGTGTAATAAATTTCCAG 263

RESULT 13
US-10-084-817-70
; Sequence 70, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Plon
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: FA-0046 US
; CURRENT APPLICATION NUMBER: US/10/084,817
; PRIOR FILING DATE: 2002-02-25
; PRIOR APPLICATION NUMBER: 60/270,784
; PRIOR FILING DATE: 2001-02-23
; NUMBER OF SEQ ID NOS: 365
; SOFTWARE: PERL Program
; SEQ ID NO 70
; LENGTH: 2905
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20030119009A1 040652.36
US-10-084-817-70

Query Match
Best Local Similarity 7.2%; Score 157; DB 14; Length 2905;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1813 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 1872
DB 107 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 166

QY 1873 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTGATTCCTGCAATTCCTGAAG 1932
DB 167 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTGATTCCTGCAATTCCTGAAG 226

QY 1933 ATCTCAAGATCTGGACTACTGTGTAATAAATTTCCAG 1969
DB 227 ATCTCAAGATCTGGACTACTGTGTAATAAATTTCCAG 263
```

```
RESULT 14
US-09-971-392-88
; Sequence 88, Application US/09971392
; Publication No. US20030134283A1
; GENERAL INFORMATION:
; APPLICANT: Peterson, David P.
; APPLICANT: Peterson, Cecelia I.
; APPLICANT: Cocks, Benjamin G.
; TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
; FILE REFERENCE: FA-0029 US
```

```
; CURRENT APPLICATION NUMBER: US/09/971,392
; CURRENT FILING DATE: 2001-10-03
; PRIOR APPLICATION NUMBER: 60/237,652
; PRIOR FILING DATE: 2000-10-03
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: PERL Program
; SEQ ID NO 88
; LENGTH: 2906
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Template ID: 040652.29
US-09-971-392-88

Query Match
Best Local Similarity 7.2%; Score 157; DB 10; Length 2906;
Matches 157; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1813 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 1872
DB 107 AGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTTTCTGGG 166

QY 1873 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTGATTCCTGCAATTCCTGAAG 1932
DB 167 AACTTTACTGATTATATCTCCCGCTCACAAATAAGCATTTGATTCCTGCAATTCCTGAAG 226

QY 1933 ATCTCAAGATCTGGACTACTGTGTAATAAATTTCCAG 1969
DB 227 ATCTCAAGATCTGGACTACTGTGTAATAAATTTCCAG 263

RESULT 15
US-10-433-544-13
; Sequence 13, Application US/10433544
; Publication No. US20040053396A1
; GENERAL INFORMATION:
; APPLICANT: JACKSON, Jennifer L.; BAUGHN, Maria R.;
; APPLICANT: YUE, Henry; GERTZEN, Kimberly J.;
; APPLICANT: TANG, Y. Tom; THANGAVELU, Kavitha;
; APPLICANT: LAL, Preeti G.; XU, Yuming;
; APPLICANT: ELLIOTT, Vicki S.; LU, Dyung Aina M.;
; APPLICANT: YAO, Monique G.; LU, Yan;
; APPLICANT: HAFALIA, April J.A.; RAMKUMAR, Jayalaxmi
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PF-0858 USN
; CURRENT APPLICATION NUMBER: US/10/433,544
; CURRENT FILING DATE: 2003-06-02
; PRIOR APPLICATION NUMBER: PCT/US01/46874
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/251,791
; PRIOR FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PERL Program
; SEQ ID NO 13
; LENGTH: 1892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20040053396A1 2344051CB1
US-10-433-544-13

Query Match
Best Local Similarity 4.9%; Score 106.4; DB 12; Length 1892;
Matches 135; Conservative 0; Mismatches 26; Indels 3; Gaps 1;

QY 1806 TCTTTGCAGATACCTTCATTTTCTTAGCGTTTCTGGAGATTACAAATCCTCGGTTCCGTT 1865
DB 2 TCTTTGAAATACCTTCATTTTCTTAGCATTTTCAGGAGATTATAACATCTCTGATTTTCTAGT 61

QY 1866 TCTTGGGAATCTTACTGATTTATCTCCCGCTCACAAATAAGCATTTGATTCCTGCAATTCCTGAAG 1925
```


us-09-931-733-1.rnpb

Mon Mar 29 09:55:01 2004

Db 62 TTCTGAGAGCTTACTGACTGATTCCCTATTCA--AAACAATCCTCATTTCTCTACATT 118
Oy 1926 TCTGAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 1969
Db 119 TCTGAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCAG 162

Search completed: March 28, 2004, 07:59:35
Job time : 796 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 27, 2004, 23:46:19 ; Search time 5729 Seconds

(without alignments)
11420.499 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 2191

Sequence: 1 tatcaaatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pin:*
20: em_ges_vrt:*
21: em_ges_fun:*
22: em_ges_nam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_phg:*
26: em_ges_vhl:*
27: em_ges_vrl:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	244	11.1	300	9 AU100150	AU100150 AU100150
2	240	11.0	480	9 AL702095	AL702095 DKFZP686A
3	240	11.0	656	13 BU927876	BU927876 AGENCOURT
4	240	11.0	742	14 CF596894	CF596894 AGENCOURT

5	239.4	10.9	497	12	BM819118
6	238.4	10.9	746	14	CD639214
7	237.4	10.8	406	10	AW139342
8	234	10.7	746	14	CD104534
9	233	10.6	608	9	AL711311
10	233	10.6	854	12	BG253409
11	232	10.6	873	14	CD252021
12	232	10.6	1014	12	BM463414
13	231.4	10.6	817	13	BU934276
14	231	10.5	714	12	BG540026
15	223	10.2	579	9	AU279766
16	221.2	10.1	647	14	CD690890
17	218.6	10.0	482	14	CB161132
18	218	9.9	472	14	W94278
19	218	9.9	939	13	EX437148
20	212	9.7	367	9	AA356204
21	212	9.7	752	10	BE888674
22	211.8	9.7	263	9	AA471173
23	209	9.5	603	9	AL702012
24	207.4	9.5	721	12	BI669261
25	207.4	9.5	817	12	BI603142
26	206.8	9.4	873	12	BG256677
27	206	9.4	354	9	AA355830
28	204.4	9.3	1016	12	BI835314
29	199.8	9.1	480	14	N31546
30	198.8	9.1	1388	13	BQ226850
31	197	9.0	402	9	AV737596
32	196.4	9.0	611	12	BI596305
33	196	8.9	592	10	AW271353
34	195.6	8.9	996	9	AL521077
35	193.6	8.8	649	12	BI561720
36	192.4	8.8	612	13	BX485700
37	191.8	8.8	1201	9	AL534190
38	179	8.2	450	9	AI830394
39	169	7.7	751	13	B0618394
40	160.8	7.3	438	9	AA115983
41	159	7.3	612	10	BF058420
42	158	7.2	729	14	CA418306
43	157	7.2	681	12	BG914201
44	157	7.2	850	10	BG030589
45	155.8	7.1	694	12	BG621446

ALIGNMENTS

RESULT 1
AU100150

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AU100150 Sugano Homo sapiens cDNA library Homo sapiens cDNA clone
LNG15174 similar to Human interferon-gamma induced protein (IFI 16)
gene, mRNA sequence.

AU100150.1 GI:13551279

EST.

Homo sapiens

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 300)

Suzuki,Y., Tsunoda,T., Taira,H., Mizushima-Sugano,J., Sese,J.,

Hata,H., Ota,T., Isega,T., Tanaka,T., Nakamura,Y., Morishita,S.,

Okubo,K., Suyama,A. and Sugano,S.

In silico mapping of the 5'-ends of human mRNAs using full-length

enriched and 5'-end enriched cDNA libraries constructed by

Oligo-capping method

Unpublished (2001)

Contact: Yutaka Suzuki

Department of Virology

Institute of Medical Science, University of Tokyo

4-6-1, Shirokanedai, Minatoku, Tokyo 108-8639, Japan

Email: yusuzuki@ims.u-tokyo.ac.jp

Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and

Sugano, S. Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. Gene 200 (1-2), 149-156 (1997).

FEATURES

source
Location/Qualifiers
1. 300
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UIG15174"
/clone_lib="Sugano Homo sapiens cDNA library"

ORIGIN

Query Match 11.1%; Score 244; DB 9; Length 300;
Best Local Similarity 99.6%; Pred. No. 3.8e-49;
Matches 255; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1715 ACTCCATTTCTCTGGGG-AAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 1773
Db 2 ACTCCATTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 61
QY 1774 AGTGACTCAACGAGCCCTTTTCCCTTGTATCTTTCAGATCTTCACTTCTTAGCG 1833
Db 62 AGTGACTCAACGAGCCCTTTTCCCTTGTATCTTTCAGATCTTCACTTCTTAGCG 121
QY 1834 TTTCTGGAGATTACAAATCTCTGCGGTTCCGTTTCTGGAACTTTTACTGATTATCTCCC 1893
Db 122 TTTCTGGAGATTACAAATCTCTGCGGTTCCGTTTCTGGAACTTTTACTGATTATCTCCC 181
QY 1894 CCCTCACACAATAAGCATTTGATTCCTGCATTTCTGAAGATCTGAAGATCTGGACTACTG 1953
Db 182 CCCTCACACAATAAGCATTTGATTCCTGCATTTCTGAAGATCTGAAGATCTGGACTACTG 241
QY 1954 TTGAAGAAATTTCCAG 1969
Db 242 TTGAAGAAATTTCCAG 257

RESULT 2

AL702095 480 bp mRNA linear EST 04-SEP-2003
LOCUS DKFP686A15155 r1 686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION DKFP686A15155 5', mRNA sequence.
ACCESSION AL702095
VERSION AL702095.1 GI:19685451
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Bloecker, H., Boecker, M., Brandt, P., Mewes, W., Weil, B. and Wiemann, S.
TITLE EST (Bloecker, H., Boecker, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.)
JOURNAL Unpublished (1999)
COMMENT Contact: MIPS
MIPS Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.
No sl sequence available.
This clone (DKFP686A15155) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clones@rzpd.de.

FEATURES

source
Location/Qualifiers
1. 480
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

/clone="DKFP686A15155"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
cDNA-collection"

ORIGIN

Query Match 11.0%; Score 240; DB 9; Length 480;
Best Local Similarity 99.6%; Pred. No. 3.9e-48;
Matches 251; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1719 CATTTTCTCTGGGG-AAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 1777
Db 2 CATTTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCTAGTCACTA 61
QY 1778 ACTCAACCAAGGCCCTTTTCCCTTGTATCTTTCAGATCTTCACTTCTTAGCGTTTC 1837
Db 62 ACTCAACCAAGGCCCTTTTCCCTTGTATCTTTCAGATCTTCACTTCTTAGCGTTTC 121
QY 1838 TGGAGATTACAAATCTCTGCGGTTCCGTTTCTGGAACTTTTACTGATTATCTCCCCCT 1897
Db 122 TGGAGATTACAAATCTCTGCGGTTCCGTTTCTGGAACTTTTACTGATTATCTCCCCCT 181
QY 1898 CACACAATAAGCATTTGATTCCTGCATTTCTGAAGATCTCAAGATCTGACTACTGTGA 1957
Db 182 CACACAATAAGCATTTGATTCCTGCATTTCTGAAGATCTCAAGATCTGACTACTGTGA 241
QY 1958 AAAAATTTCCAG 1969
Db 242 AAAAATTTCCAG 253

RESULT 3

BU927876 656 bp mRNA linear EST 18-OCT-2002
LOCUS AGENCOURT 10434296 NIH MGC 126 Homo sapiens cDNA clone
DEFINITION IMAGE:6653388 5', mRNA sequence.
ACCESSION BU927876
VERSION BU927876.1 GI:24116606
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC http://mgs.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: NCI
cDNA Library Preparation: Michael Brownstein Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LICM2903 row: f column: 12
High quality sequence stop: 555.
Location/Qualifiers
1. 656
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6653388"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH MGC 126"
/note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgctcgccc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -

FEATURES

source
Location/Qualifiers
1. 656
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6653388"
/tissue_type="mixed (pool of 40 RNAs)"
/lab_host="DH10B (T1-phage-resistant)"
/clone_lib="NIH MGC 126"
/note="Vector: pDNR-LIB; Site_1: SfiI (ggccattatggcc); Site_2: SfiI (ggccgctcgccc); Double-stranded cDNA was prepared from a pool of 40 cell line polyA+ RNAs (bladder - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -

4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%, kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%, ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were used in cloning as follows:
 5'-AAGCAGTGGTATCAAGCAGGCGGCGG-3' and 5'-ATTCTAGAGCCGAGCGGCGGCGG-3'. Full-length enriched library was constructed using the Clontech Creator SMART kit and size-selected to contain the 0.5-1 kb size fraction (other fractions present in NIH MGC 127 and NIH MGC 128). Library created in the laboratory of T. Uesdin, M.D., Ph.D. (NIH, NIH). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 11.0%; Score 240; DB 13; Length 656;
 Best Local Similarity 99.6%; Pred. No. 4e-48;
 Matches 251; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1719 CATTTTCTCTGGGG-AATAGCAGAAATAGGAGCAGCAGCAGCTAGTCACTAAGTG 1777
 Db 4 CATTTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCAGCTAGTCACTAAGTG 63
 QY 1778 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTGCGAGATCTTCATTTCTTAGCGTTTC 1837
 Db 64 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTGCGAGATCTTCATTTCTTAGCGTTTC 123
 QY 1838 TGGAGATTACCAATCTCGGGTTCCTGTTCTTGGGAATCTTACTGATTTATCTCCCGCTT 1897
 Db 124 TGGAGATTACCAATCTCGGGTTCCTGTTCTTGGGAATCTTACTGATTTATCTCCCGCTT 183
 QY 1898 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGAGACTACTGTGA 1957
 Db 184 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGAGACTACTGTGA 243
 QY 1958 AAAAATTTCCAG 1969
 Db 244 AAAAATTTCCAG 255

RESULT 4

CF596894 742 bp mRNA linear EST 26-SEP-2003
 LOCUS AGENCOURT 15667536 NICHDS Hs Ovl Homo sapiens cDNA clone
 DEFINITION IMAGE:30704931 5', mRNA sequence.
 CF596894
 CF596894.1 GI:36353819
 EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 742)
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Daniela S. Gerhard, Ph.D.
 Office of Cancer Genomics
 National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Gregory F. Erickson, Ph.D.
 cDNA Library Preparation: Invitrogen Corp
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDCM254 row: m column: 04
 High quality sequence stop: 498.
 Location/Qualifiers
 1. .742
 /organism="Homo sapiens"

FEATURES

source

/mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30704931"
 /tissue_type="Ovary"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NICHDS Hs Ovl"
 /note="Vector: pDNR-f1b; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggccgctcgcc); Library is oligo-dT primed and directionally cloned. Granulosa lutein cells aspirated from preovulatory follicles of normal cycling women undergoing ovulation induction for infertility due to male factor and normal donors. The cells were from follicles stimulated with Lupron, FSH and hCG. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CAGGCGCATATGGCC-3' and 3' adaptor sequence: 5'-ATTCTAGAGCGGCGGCGGCGGCGG-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 2.23 kb (range 1.0-4.5 kb). 14/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."

ORIGIN

Query Match 11.0%; Score 240; DB 14; Length 742;
 Best Local Similarity 99.6%; Pred. No. 4e-48;
 Matches 251; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
 QY 1719 CATTTTCTCTGGGG-AATAGCAGAAATAGGAGCAGCAGCAGCTAGTCACTAAGTG 1777
 Db 5 CATTTTCTCTGGGGCAATAGCAGAAATAGGAGCAGCAGCAGCTAGTCACTAAGTG 64
 QY 1778 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTGCGAGATCTTCATTTCTTAGCGTTTC 1837
 Db 65 ACTCAACCAAGGCGCTTTTCTTGTTATCTTTGCGAGATCTTCATTTCTTAGCGTTTC 124
 QY 1838 TGGAGATTACCAATCTCGGGTTCCTGTTCTTGGGAATCTTACTGATTTATCTCCCGCTT 1897
 Db 125 TGGAGATTACCAATCTCGGGTTCCTGTTCTTGGGAATCTTACTGATTTATCTCCCGCTT 184
 QY 1898 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGAGACTACTGTGA 1957
 Db 185 CACACAATAAGCATGATTCCTGATTTCTGAAGATCTCAAGATCTGAGACTACTGTGA 244
 QY 1958 AAAAATTTCCAG 1969
 Db 245 AAAAATTTCCAG 256

RESULT 5

BM819118 497 bp mRNA linear EST 06-MAR-2002
 LOCUS K-EST0086922 S11SNUI Homo sapiens cDNA clone S11SNUI-28-C10 5',
 DEFINITION mRNA sequence.
 ACCESSION BM819118
 VERSION BM819118.1 GI:19175531
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 497)
 AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R., Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and Kim, Y.S.
 TITLE 21C Frontier Korean EST Project 2001
 JOURNAL Unpublished (2002)
 COMMENT Contact: Kim YS
 Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr
 Plate: 28 row: C column: 10

High quality sequence stop: 497.

FEATURES

Location/Qualifiers
 1. 497
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="S11SNU1-28-C10"
 /sex="M"
 /tissue_type="Stomach"
 /cell_type="Lymphoblast-like"
 /cell_line="SNU-1"
 /lab_host="Top10F"
 /clone_lib="S11SNU1"

/note="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI; Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dT primer. After first strand synthesis, RNA was degraded by NAC treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME18S-PL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

ORIGIN

Query Match 10.9%; Score 239.4; DB 12; Length 497;

Best Local Similarity 99.6%; Pred. No. 5.4e-48;

Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1729 GGGGAATAGCAGATAGGAGCAAGCAGCAGCTAGTCAGCTACTACTAGTCACTCAACCAAG 1788
 DB 2 GGGCAATAGCAGATAGGAGCAAGCAGCAGCTAGTCAGCTACTACTAGTCACTCAACCAAG 61
 QY 1789 GCCTTTTCCCTGTTTATCTTTCGAGATCTTCATTTCTTAGCGTTTCGAGATTACA 1848
 DB 62 GCCTTTTCCCTGTTTATCTTTCGAGATCTTCATTTCTTAGCGTTTCGAGATTACA 121
 QY 1849 ACATCTCGGTTCCGTTTCGGGACCTTACTGATTTATCTCCCGCTCACACAAATAA 1908
 DB 122 ACATCTCGGTTCCGTTTCGGGACCTTACTGATTTATCTCCCGCTCACACAAATAA 181
 QY 1909 GCATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTCTGTGAAAAAATTTCCA 1968
 DB 182 GCATTGATTCCTGCAATTTCTGAAGATCTCAAGATCTGGACTCTGTGAAAAAATTTCCA 241
 QY 1969 G 1969
 DB 242 G 242

RESULT 6

CD639214

LOCUS 746 bp mRNA linear EST 17-JUN-2003
 DEFINITION AGENCOURT_14533326 NIH_MGC_191 Homo sapiens cDNA clone
 IMAGE:30414306 5', mRNA sequence.

ACCESSION CD639214

VERSION CD639214.1 GI:31805803

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 746)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

CONTACT: Daniela S. Gerhard, Ph.D.

COMMENT Office of Cancer Genomics

National Cancer Institute / NIH
 Bldg. 31 Rm10A07 Bethesda, MD 20892
 Email: cgabs-r@mail.nih.gov

Tissue Procurement: Narayan Bhat

CDNA Library Preparation: CLONTECH Laboratories, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: NDCM210 row: k column: 19

High quality sequence stop: 523.

FEATURES

source

Location/Qualifiers
 1. 746
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30414306"
 /tissue_type="Pooled"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_lib="NIH_MGC_191"
 /note="Vector: pNCR-Lib; Site 1: SfiI (ggccattatggcc); Site 2: SfiI (ggcgctctggcc); Library is oligo-dT primed and directionally cloned. PBMC - Peripheral Blood Mononuclear Cells. RNA was pooled from 3/6hour stimulation with PMA adn Ionomycin. 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCAATAGGCC-3' and 3' adaptor sequence: 5'-ATTAGAGGCGGCGGCGGCGCATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.69 kb (range 0.70-5.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH_MGC Library."

ORIGIN

Query Match 10.9%; Score 238.4; DB 14; Length 746;

Best Local Similarity 99.2%; Pred. No. 1e-47;

Matches 250; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1719 CATTTTCTCTGGG-AAATAGCAGATAGGAGCAAGCCAGCAGCTAGTCAGTAACTAAGTG 1777
 DB 4 CATTTTCTCTGGGCAATAGCGAATAGGAGCAAGCCAGCAGCTAGTCAGTAACTAAGTG 63
 QY 1778 ACTCAACCAAGCGCTTTTCTCTGTTATCTTCAGATACCTTCATTTCTTAGGTTTC 1837
 DB 64 ACTCAACCAAGCGCTTTTCTCTGTTATCTTCAGATACCTTCATTTCTTAGGTTTC 123
 QY 1838 TGGAGATTACAACATCTCTCGGTTCCGTTTCTGGGAACCTTTACTGATTTATCTCCCGCT 1897
 DB 124 TGGAGATTACAACATCTCTCGGTTCCGTTTCTGGGAACCTTTACTGATTTATCTCCCGCT 183
 QY 1898 CACACAAATAAGCATTTGATTCCTGATTTCTGAAGATCTCAAGATCTGAGATCTGTGA 1957
 DB 184 CACACAAATAAGCATTTGATTCCTGATTTCTGAAGATCTCAAGATCTGAGATCTGTGA 243
 QY 1958 AAAAATTTCCAG 1969
 DB 244 AAAAATTTCCAG 255

RESULT 7

AW139342

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Homo sapiens

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

406 bp mRNA linear EST 30-OCT-1999
 UI-H-BRI-ada-f-06-0-UI.s1 NCI_CGAP_Sub3 Homo sapiens cDNA clone
 IMAGE:2716186 3', mRNA sequence.

AW139342

AW139342.1 GI:6144060

EST.

Homo sapiens (human)

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 406)
 AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA library Preparation: M.B. Soares Lab Clone Distribution:
 NCI-CGAP clone distribution information can be found through the
 I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES Location/Qualifiers
 source 1..406
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2716186"
 /lab_host="DH10B (Life Technologies)"
 /clone_lib="NCI CGAP Sub3"
 /notes="Vector: p773D-Pac (Pharmacia) with a modified
 polylinker. Site 1: Not 1; Site 2: Eco RI; The
 NCI CGAP Sub3 library is a subtracted library derived from
 the NCI CGAP Sub1 library, which is a subtracted library
 derived from B1. B1 constitutes a mixture of 21
 normalized or subtracted NCI CGAP libraries:
 NCI CGAP Co4, NCI CGAP Pr22, NCI CGAP Pr28, NCI CGAP Co10,
 NCI CGAP Co16, NCI CGAP Kid5, NCI CGAP Kid12,
 NCI CGAP Kid3, NCI CGAP Kid11, NCI CGAP Lym2,
 NCI CGAP Br2, NCI CGAP Co8, NCI CGAP CLL1, NCI CGAP Lei2,
 NCI CGAP Brn23, NCI CGAP Lu5, NCI CGAP Lu24,
 NCI CGAP Lu19, NCI CGAP GC4, NCI CGAP GC6,
 NCI CGAP Brn25. These 21 libraries were pooled and a
 single-stranded DNA preparation of the resulting mixture
 was used as a tracer in a subtractive hybridization with
 a driver whose composition is detailed below:
 NCI CGAP Kid3 pool 1 LLAM 3334-3337, 3682-3683,
 3798-3803 (IMAGE CloneIDs 1322376-1323911,
 1456008-1456775, 1500552-1502855); NCI CGAP Kid5 pool 1
 LLAM 3338-3342, 3722-3725, 3776-3778 (IMAGE CloneIDs
 1323912-1325831, 1471368-1472903, 1492104-1493255);
 NCI CGAP Lu5 pool 1 LLAM 3575-3582, 3651-3654 (IMAGE
 CloneIDs 1414920-1417991, 1520904-1522439); NCI CGAP GC4
 pool 1 LLAM 3164-3167, 3716-3720, 3733-3735 (IMAGE
 CloneIDs 1257096-1258631, 1469064-1470983,
 1475592-1476743); NCI CGAP Pr22 pool 1 LLAM 2457-2459,
 2758-2759, 3062-3068 (IMAGE CloneIDs 985608-986759,
 1101192-1101959, 1217928-1220615); NCI CGAP Co10 pool 1
 LLAM 2644-2653, 2871-2872 (IMAGE CloneIDs 1057416-1061255,
 1145844-1145351). Subtraction was performed as previously
 described [Bonaldo, Lennon & Soares (1996): Normalization
 and Subtraction: Two Approaches To Facilitate Gene
 Discovery. Genome Research 6, 791-806.
 TAG_TISSUE=Lymph node
 TAG_LIB=NCI CGAP_Lym2
 TAG_SEQ=AAATG"

ORIGIN
 Query Match 10.8%; Score 237.4; DB 10; Length 406;
 Best Local Similarity 98.8%; Pred. No. 1.7e-47;
 Matches 238; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1729 GGGGAATAGCAGATAGGACGACGACCTAGTCAGCTAACTAGTCACTCAACCAAG 1788
 Db 16 GGGCAATAGCAGATAGGACGACGACCTAGTCAGCTAACTAGTCACTCAACCAAG 75
 QY 1789 GCGTTTTTCTTGTATCTTTCAGATACCTTCATTTTCTTAGCGTTCTCGAGATACA 1848
 Db 76 GCGTTTTTCTTGTATCTTTCAGATACCTTCATTTTCTTAGCGTTCTCGAGATACA 135

QY 1849 ACATCTCGGGTTCGGTTCCTGGAACTTTACTGATTATCTCCCTCACAACAATAA 1908
 Db 136 ACATCTCGGGTTCGGTTCCTGGAACTTTACTGATTATCTCCCTCACAACAATAA 195
 QY 1909 GCATTGATTCCTGCATTTCTTGAAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCA 1968
 Db 196 GCATTGATTCCTGCATTTCTTGAAGATCTCAAGATCTGGACTACTGTGAAAAAATTTCCA 255
 QY 1969 G 1969
 Db 256 G 256

RESULT 8
 CD104534 746 bp mRNA linear EST 15-MAY-2003
 LOCUS AGENCOURT 13979535 NIH_MGC 186 Homo sapiens cDNA clone
 DEFINITION IMAGE:30371137 5', mRNA sequence.
 ACCESSION CD104534
 VERSION CD104534.1 GI:30757708
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 746)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Dr. Michael Brownstein and Dr. Miklos Palkovits
 cDNA Library Preparation: CLONTECH Laboratories, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: NDCM161 row: e column: 02
 High quality sequence stop: 561.

FEATURES Location/Qualifiers
 source 1..746
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:30371137"
 /lab_host="DH10B (TI phage-resistant)"
 /clone_lib="NIH MGC 186"
 /note="Organ: Pooled Skin; Vector: pDNR-LTB; Site 1: SfiI
 (ggccattagggcc); Site 2: SfiI (ggccgcttcggcc); Library is
 oligo-dT primed and directionally cloned. cDNA was
 prepared from a pooled samples of tissues from skin,
 meninges, duramater, pia matter and choroid plexus. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-CACGCCATTATGGC-3' and 3' adaptor
 sequence: 5'-ATTCAGAGCGCGGCGCCGACATG-dT(30)BN-3'
 (where B = A, C, or G and N = A, C, G, or T). Average
 insert size 1.47 kb (range 0.50-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA). Note: this is a NIH_MGC
 Library"

ORIGIN
 Query Match 10.7%; Score 234; DB 14; Length 746;
 Best Local Similarity 100.0%; Pred. No. 1.2e-46;
 Matches 234; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1736 AGCAGATAGGACGACGACCTAGTCAGCTAACTAGTCACTCAACCAAGGCCCTTT 1795
 Db 3 AGCAGATAGGACGACGACCTAGTCAGCTAACTAGTCACTCAACCAAGGCCCTTT 62

Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA Project, Sugiyama, T.; Wakamatsu, A.; Irie, R.;
Umezawa, A.; Fukuma, M.; Kuwakari, S.; Hata, J.; Ishii, S.; Yamamoto, J.;
Isono, Y.; Saito, K.; Nakamura, Y.; Masuho, Y.; Nagai, K.; Isogai, T.
HRI human cDNA project; cDNA library construction & 5'-end one
pass sequencing; Helix Research Institute.

FEATURES

source
1..579
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CHONS2001343"
/cell_type="chondrocytes"
/clone_lib="CHONS2"
/note="Vector: pME18SFL3"

ORIGIN

Query Match 10.2%; Score 223; DB 9; Length 579;
Best Local Similarity 99.6%; Pred. No. 6.1e-44;
Matches 234; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 1736 AGCAGAAATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTGAAGTCAACCAAGGCCCTTTT 1795
Db 1 AGCAGAAATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTGAAGTCAACCAAGGCCCTTTT 60
QY 1796 TTCTCTGTTATCTTTGAGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCT 1855
Db 61 TTCTCTGTTATCTTTGAGATACCTTCATTTCTTAGCGTTTCTGGAGATTACAAATCCT 120
QY 1856 GCGGTTCCGTTTCTGGGAACCTTTAC-TGATTATCTCCCGCTCACACAAATAGCATTG 1914
Db 121 GCGGTTCCGTTTCTGGGAACCTTTACNTGATTTATCTCCCGCTCACACAAATAGCATTG 180
QY 1915 ATTCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 1969
Db 181 ATTCTGCAATTTCTGAAGATCTCAAGATCTGGACTACTGTTGAAAAAATTTCCAG 235

Search completed: March 28, 2004, 03:14:27
Job time : 5734 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 05:37:57 ; Search time 158 Seconds

(without alignments)

7836.223 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p_model -DEV=xlh
-Q=/cgn2_1/USPRO.spool/US09931733/runat 25032004 155916 9625/app_query.fasta_1.2375
-DB=A_Geneseq 29Jan04 -QMT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blotsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733 @CGN 1 1 167 @runat 25032004 155916 9625 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq 29Jan04:*

1: geneseqp1980s:.*
2: geneseqp1990s:.*
3: geneseqp2000s:.*
4: geneseqp2001s:.*
5: geneseqp2002s:.*
6: geneseqp2003as:.*
7: geneseqp2003bs:.*
8: geneseqp2004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95.5	2.5	203	AAU19374	AAU19374 Human G p
2	88	2.3	128	AAU99347	AAU99347 Human epi
3	87	2.3	147	AAU99349	AAU99349 Human epi
4	87	2.3	322	ABP43964	ABP43964 mRNA for
5	87	2.3	331	ABP43964	ABP43964 mRNA for
6	86.5	2.3	944	ABP39507	ABP39507 DNA polym
7	85.5	2.2	331	AAV90653	AAV90653 Human mut
8	85.5	2.2	331	ADC22703	ADC22703 Human G p
9	85.5	2.2	456	ADA33128	ADA33128 Acinetoba
10	83	2.2	331	AAV44486	AAV44486 Human GPR

C 11	83	2.2	331	4	AAU78390	Human pro
C 12	83	2.2	331	4	AAU38949	Human pol
C 13	83	2.2	331	6	ABP81884	Human G p
C 14	83	2.2	333	4	AAU82048	Human G p
C 15	83	2.2	340	4	AAU79374	Human pro
16	82.5	2.2	310	4	AAU73008	Human pro
17	82.5	2.2	310	4	AAU72789	Human oif
18	82.5	2.2	310	7	ADD47232	Rat Prote
19	81.5	2.1	582	4	AAU03896	G protein
C 20	81	2.1	34	4	AAU17710	Novel hum
C 21	80.5	2.1	447	6	ABP57694	Saccharop
22	80	2.1	260	2	AAU27871	Odorant r
23	80	2.1	312	2	AAU27873	Odorant r
24	80	2.1	312	4	AAU73009	Olfactory
C 25	80	2.1	331	3	AAU90619	Human G p
C 26	80	2.1	331	7	ADC22555	Human G p
27	79.5	2.1	272	3	AAU84944	A human c
28	79.5	2.1	272	3	AAU97289	Lipid ass
C 29	79.5	2.1	330	5	ABP28451	Streptoco
30	79	2.1	362	7	ADC87555	Human GPC
31	78	2.0	110	7	ADB36332	Human imm
32	78	2.0	118	7	ADB36331	Human imm
33	78	2.0	139	7	ADB36333	Human imm
34	78	2.0	212	5	AAU47347	Human Pax
35	78	2.0	252	6	ABU57655	Different
36	78	2.0	377	2	AAU48693	G-protein
37	78	2.0	377	2	AAU02665	G-protein
C 38	78	2.0	409	2	AAU03544	Bacillus
C 39	78	2.0	432	3	AAU05849	Arabidops
C 40	77.5	2.0	314	6	ABM69332	Photornab
41	77.5	2.0	475	6	ABU38376	Protein e
42	77.5	2.0	532	4	ABG10280	Novel hum
43	77.5	2.0	631	4	AAU93865	Human pol
44	77	2.0	329	4	ABG16890	Novel hum
C 45	77	2.0	1105	4	AAU40174	Human pol

ALIGNMENTS

RESULT 1

AAU19374
ID AAU19374 standard; protein; 203 AA.
XX

AAU19374;

AC (first entry)

04-DEC-2001

Human G protein-coupled receptor nGPCR-2307.

Human; G protein-coupled receptor; nGPCR-x; antiviral; analgesic; cytostatic; cardiac; antidiabetic; anorectic; hypotensive; hypertensive; antiparkinsonian; nootropic; neuroprotective; antidepressant; viral infection; HIV-1; human immunodeficiency virus; HIV-2; pair; cancer; metabolic disease; cardiovascular disease; type 2 diabetes; obesity; anorexia; hypotension; hypertension; myocardial infarction; atherosclerosis; Parkinson's disease; psychosis; neurological disorder; schizophrenia; migraine; major depression; anxiety; mental disorder; manic depression; dyskinesia; Huntington's disease; Tourette's Syndrome.

XX Homo sapiens.

XX WO200166751-A2.

XX 13-SEP-2001.

PD 08-MAR-2001; 2001WO-US0007370.

XX 08-MAR-2000; 2000US-0187583P.

PR 08-MAR-2000; 2000US-0187584P.

PR 08-MAR-2000; 2000US-0187637P.

PR 08-MAR-2000; 2000US-0187639P.

PR 08-MAR-2000; 2000US-0187640P.

PR 08-MAR-2000; 2000US-0187707P.

08-MAR-2000; 2000US-0187708P.
 08-MAR-2000; 2000US-0187709P.
 08-MAR-2000; 2000US-0187827P.
 08-MAR-2000; 2000US-0188290P.
 08-MAR-2000; 2000US-0188292P.
 08-MAR-2000; 2000US-0188293P.
 (PHAA) PHARMACIA & UPJOHN CO.
 Vogeli G;
 WPI; 2001-536779/59.
 N-PSDB; AAS30946.
 Isolated nucleic acid molecules encoding G protein-coupled receptors termed nGPCR-x, useful in the treatment and diagnosis of viral infections, cancers and mental disorders (e.g. Parkinson's disease and schizophrenia).
 Claim 31; Page 267; 292pp; English.
 The invention relates to novel isolated nucleic acid molecules encoding G protein-coupled receptors termed nGPCR-x. nGPCR-x polynucleotides, polypeptides, and modulators may be used in the treatment of diseases and conditions such as infections, such as viral infections caused by HIV-1 (human immunodeficiency virus) or HIV-2, pain, cancers, metabolic and cardiovascular diseases and disorders (e.g., type 2 diabetes, obesity, anorexia, hypotension, hypertension, myocardial infarction, atherosclerosis), Parkinson's disease, and psychotic and neurological disorders, including schizophrenia, migraine, major depression, anxiety, mental disorder, manic depression, and dyskinesias, such as Huntington's disease or Tourette's Syndrome and many other diseases and syndromes listed in the specification. nGPCR-x polynucleotides and polypeptides, as well as nGPCR-x modulators, may also be used in diagnostic assays for such diseases or conditions. The present sequence represents a G protein-coupled receptor of the invention
 SQ Sequence 203 AA;
 Alignment Scores:
 Pred. No.: 0.0366 Length: 203
 Score: 95.50 Matches: 34
 Percent Similarity: 36.36% Conservative: 14
 Best Local Similarity: 25.76% Mismatches: 41
 Query Match: 2.50% Indels: 43
 DB: 4 Gaps: 4
 US-09-931-733-1 (1-2191) x AAU19374 (1-203)
 QY 1329 CATGTTGCTCTCTTATATTGAGCATCATTTA----- 1361
 Db 91 HisSerAlaLeuMetLeuPheHisValCysLeuPheThrValProgluLeuCys 110
 QY 1362 -----AGGATATTAAGTCTCTTAATG 1382
 Db 111 ValCysLeuLeuLeuLeuTyArgIleHisSerLeuLeuLeuLeuAspLeuPhe 130
 QY 1383 ACTCAGTAACTAGTCTGGGTAGAGAGTCTTAATGTCCTTCCTGATGCTACG 1442
 Db 131 Thr-----Ser 132
 QY 1443 AATTGTTAGAAAACAACTATTGTTTATATACATTATGATACAAAGTGTAAAT 1502
 Db 133 AsnThrIleTrpGlyPheSerIleCys-----IleAsnLeuArgGlyThrHisAlaVal 150
 QY 1503 TCGTTACATGGAATATATGCAATTGCTGAAGTTCAGAGATTCAGTGCATACACCG 1562
 Db 151 PheLeuHisArgTyAsnAlaGluTrpArgSerLeuLeuLeuLeuLeuLeuLeu 170
 QY 1563 AAAATGTTAACTGATCCCAT---TAAAGTAAATTTCTCATCCCATTTCCCTCACCAC 1619
 Db 171 AsnAsnValHisCysThrHisValAsnSerHisProSerAsnProLeuProSerHis 190

QY 1620 ATGCTCAGCCCTGTCAGGAAACTGTTCTATTTCT 1655
 Db 191 ProProSerPheLeuAlaSerAsnValTy-Tyr-Ser 202
 RESULT 2
 AAU99347
 ID AAU99347 standard; protein; 128 AA.
 AC AAU99347;
 DT 07-OCT-2002 (first entry)
 DE Human epididymal secretory protein E3 alpha precursor (CCS3, Q14507).
 KW Human; vaccine; gene therapy; cytokine; angiogenin; 1B1E:A;
 KW Biopendium search database; CCS3; Q14507;
 KW epididymal secretory protein E3 alpha precursor; immune response;
 KW cell growth; inflammation; wound healing; embryogenesis; apoptosis;
 KW oncology; cancer; reproductive health-related condition.
 OS Homo sapiens.
 XX WO200244382-A1.
 XX 06-JUN-2002.
 XX 28-NOV-2001; 2001WO-GB005245.
 XX 28-NOV-2000; 2000GB-00028971.
 XX (INPH-) INPHARMATICA LTD.
 XX Gutteridge A, Fagan RJ, Phelps CB;
 XX WPI; 2002-537454/57.
 XX N-PSDB; ABK87961.
 Novel isolated CCS3, CCS9 or CCS3a polypeptides which function as cytokines, and polynucleotides encoding the polypeptides, useful as diagnostic reagents for diagnosing disease related to reproductive health in patient.
 Claim 1; Fig 7; 9pp; English.
 The invention discloses three polypeptides which have structures which suggests they can function as cytokines. The polypeptides were identified using the Biopendium search database tool using the human archetypal cytokine family member, angiogenin (PDB code 1B1E:A) structure. The polypeptides identified comprise the epididymal secretory protein E3 alpha precursor (CCS3, Q14507), the epididymal secretory protein E3 precursor (CCS9, CAC17141.1) and the epididymal secretory protein E3 alpha (CCS3a, CAA53971.2). Cytokines are small secreted messenger proteins that act as potential regulators capable of effecting cellular processes, especially the immune response and cell growth, as well as inflammation, wound healing, embryogenesis and development and apoptosis. Cytokines can be used clinically as regulators of the immune system (e.g. in oncology to promote a response against thyroid cancer). The polypeptides are, therefore, useful in therapy and diagnosis of disease, for manufacturing a medicament for treating a reproductive health-related condition, for treating a disease in a patient (e.g. diseases in which there is an aberrant expression of the natural gene or the activity of the polypeptide, where the polypeptide, nucleic acid molecule, vector, ligand, compound or composition administered to the patient is an agonist (or antagonist) and as a cytokine. The polypeptide and nucleic acid molecules are also useful for identifying a compound that is effective in the treatment and/or diagnosis of a disease, for chromosome localisation, tissue localisation and for detecting differences in chromosomal location due to e.g. translocation or inversion. The pharmaceutical compositions comprising the polypeptides and/or polynucleotides are useful either as therapeutic or diagnostic compositions, as vaccines and in gene therapy. The sequence presented is the human epididymal secretory protein E3 alpha precursor (CCS3, Q14507)

```
SQ Sequence 128 AA;
Alignment Scores:
Pred. No.: 0.262 Length: 128
Score: 88.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.30% Indels: 33
DB: Gaps: 5

US-09-931-733-1 (1-2191) x AAU99347 (1-128)
QY 230 TCATCTATGAAATTCATCTTTGTGTGAATATGAGCATTCCTGAGT----- 283
Db 3 SerSerLeuLysile-----TrrpGlyileLeuLeuAlaLeuLeu 15
QY 284 -----TGGCGCTATAGTGACCGCAGTTGTGTGTATGCT----- 319
Db 16 CysileLeuCysArgLeu-----CysValTyrSerAsnAsnilleTyrTrp 30
QY 320 -----AGTCCCAATTGACACATTACTGTCATATAAAATG 352
Db 31 ArgGluPheileLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
QY 353 ATTATAAGTAAAGGAGAGAAATGTTCTCAAGGGTCATAAATCCCAAGGGTCTCTTACC 412
Db 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-ile-- 69
QY 413 CTTTGACCTCTGAGTTGGGTTCCAGTGAGTGCCTTAAT 452
Db 70 -TyrSerLeuTrpPheLysileGlnArgAlaCysilleAsn 82

RESULT 3
AAU99349
ID AAU99349 standard; protein; 147 AA.
XX
AC AAU99349;
XX
DT 07-OCT-2002 (first entry)
XX
DE Human epididymal secretory protein E3 alpha (CCS3a, CAA53971.2).
XX
KW Human; vaccine; gene therapy; cytokine; angiogenin; lBIE:A;
KW Biopendium search database; CCS3a; CAA53971.2;
KW epididymal secretory protein E3 alpha; immune response; cell growth;
KW inflammation; wound healing; embryogenesis; apoptosis; oncology; cancer;
KW reproductive health-related condition.
XX
OS Homo sapiens.
XX
PN WO200244382-A1.
XX
PD 06-JUN-2002.
XX
PF 28-NOV-2001; 2001WO-GB005245.
XX
PR 28-NOV-2000; 2000GB-00028971.
XX
PA (INPH-) INPHARMATICA LTD.
XX
PI Gutteridge A, Fagan RJ, Phelps CB;
XX
WP1; 2002-537454/57.
XX
N-PSDB; ABK37963.
XX
PT Novel isolated CCS3, CCS9 or CCS3a polypeptides which function as
PT cytokines, and polynucleotides encoding the polypeptides, useful as
PT diagnostic reagents for diagnosing disease related to reproductive health
PT in patient.
XX
PS Claim 1; Fig 15; 99pp; English.
XX
XX The invention discloses three polypeptides which have structures which
```

```
CC suggests they can function as cytokines. The polypeptides were identified
CC using the Biopendium search database tool using the human archetypal
CC cytokine family member, angiogenin (PDB code lBIE:A) structure. The
CC polypeptides identified comprise the epididymal secretory protein E3
CC alpha precursor (CCS9, Q14507), the epididymal secretory protein E3 beta
CC precursor (CCS9, CAC17141.1) and the epididymal secretory protein E3
CC alpha (CCS3a, CAA53971.2). Cytokines are small secreted messenger
CC proteins that act as potential regulators capable of effecting cellular
CC processes, especially the immune response and cell growth, as well as
CC inflammation, wound healing, embryogenesis and development and apoptosis.
CC Cytokines can be used clinically as regulators of the immune system (e.g.
CC in oncology to promote a response against thyroid cancer). The
CC polypeptides are, therefore, useful in therapy and diagnosis of disease,
CC for manufacturing a medicament for treating a reproductive health-related
CC condition, for treating a disease in a patient (e.g. diseases in which
CC there is an aberrant expression of the natural gene or the activity of
CC the polypeptide, where the polypeptide, nucleic acid molecule, vector,
CC ligand, compound or composition administered to the patient is an agonist
CC (or antagonist) and as a cytokine. The polypeptide and nucleic acid
CC molecules are also useful for identifying a compound that is effective in
CC the treatment and/or diagnosis of a disease, for chromosome localisation,
CC tissue localisation and for detecting differences in chromosomal location
CC due to e.g. translocation or inversion. The pharmaceutical compositions
CC comprising the polypeptides an/or polynucleotides are useful either as
CC therapeutic or diagnostic compositions, as vaccines and in gene therapy.
CC The sequence presented is the human epididymal secretory protein E3 alpha
CC (CCS3a, CAA53971.2)
XX
SQ Sequence 147 AA;

Alignment Scores:
Pred. No.: 0.378 Length: 147
Score: 87.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.27% Indels: 33
DB: Gaps: 5

US-09-931-733-1 (1-2191) x AAU99349 (1-147)
QY 230 TCATCTATGAAATTCATCTTTGTGTGAATATGAGCATTCCTGAGT----- 283
Db 3 SerSerLeuLysile-----TrrpGlyileLeuLeuAlaLeuLeu 15
QY 284 -----TGGCGCTATAGTGACCGCAGTTGTGTGTATGCT----- 319
Db 16 CysileLeuCysArgLeu-----CysValTyrSerAsnAsnilleTyrTrp 30
QY 320 -----AGTCCCAATTGACACATTACTGTCATATAAAATG 352
Db 31 ArgGluPheileLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
QY 353 ATTATAAGTAAAGGAGAGAAATGTTCTCAAGGGTCATAAATCCCAAGGGTCTCTTACC 412
Db 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisMetPhe-ile-- 69
QY 413 CTTTGACCTCTGAGTTGGGTTCCAGTGAGTGCCTTAAT 452
Db 70 -TyrSerLeuTrpPheLysileGlnArgAlaCysilleAsn 82

RESULT 4
ABP43964
ID ABP43964 standard; protein; 322 AA.
XX
AC ABP43964;
XX
DT 26-FEB-2003 (first entry)
XX
DE mRNA for HB3 alpha.
XX
KW Neuroprotective; immunomodulator; cancer; chromosome 14; cytostatic;
KW anti-inflammatory; gene therapy; nutritional supplement; wound; burn;
KW ulcer; Alzheimer's disease; Huntington's disease;
```

KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
vulnerary.

XX Homo sapiens.

XX WO200231111-A2.

XX 18-APR-2002.

XX 11-OCT-2001; 2001WO-US027760.

XX 12-OCT-2000; 2000US-00687527.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-426278/45.

XX N-PSDB; ABQ61208.

XX New polypeptides and their encoded proteins, useful as nutritional
PT sources or supplements, or in gene therapy, particularly for treating
PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
PT inflammation.

XX Claim 20; SEQ ID # 867; 357pp + Sequence Listing; English.

XX The invention relates to 446 newly isolated polynucleotide sequences. The
CC activity of polynucleotides of the invention may be described as,
CC vulnerary, neuroprotective, immunomodulator, cytostatic and anti-
CC inflammatory. Compositions comprising nucleic acids of the invention are
CC useful for treating a mammalian subject, or as nutritional sources or
CC supplements. These are useful in gene therapy, particularly for treating
CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
CC inflammation. The nucleic acids and polypeptides are also useful in
CC diagnostic and research methods. The sequences given in records ABP43544-
CC ABP43989 represent polypeptides encoded by polynucleotides of the
CC invention. NOTE: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences

XX SQ Sequence 322 AA;

Alignment Scores:
Pred. No.: 0.565 Length: 322
Score: 87.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.27% Indels: 33
DB: 5 Gaps: 5

US-09-931-733-1 (1-2191) x ABP43964 (1-322)

QY 230 TCATCTATGAATTCATCTTGTGGTATGAATATGAGCATTTGCTGAGT----- 283

Db 178 SerSerLeuLysIle-----TrpGlyIleLeuLeuAlaLeuLeu 190

QY 284 -----TGCGGCTATAGGTGACCGCAGTTTGTGTATGCT----- 319

Db 191 CysIleLeuCysArgLeu-----CysValTyrSerAsnAsnIleTyrTrp 205

QY 320 -----AGTCCCAATGAGCATTTACTGCATATAAATG 352

Db 206 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 225

QY 353 ATTATAAGTAAAGGCAGCAAAATGTTCTGAAGGTCATACTTCCAAGGGTTCTTACC 412

Db 226 AspValMetArgGluLysGluAlaLeuLysGlyLysSerPheHisMetPhe-Ile-- 244

QY 413 CTTTGACCTCTCGATTGGGGTTCCCGAGTGCCTTAAT 452

Db 245 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 257

RESULT 5

ABB84727

ID ABB84727 standard; protein; 331 AA.

XX ABB84727,

XX 16-MAY-2002 (first entry)

XX DNA polymerase III holoenzyme delta subunit related protein SEQ ID NO:12.
XX DNA polymerase III holoenzyme delta subunit; DNA polymerase III; DnaX;
KW hola; holB; antibacterial; tuberculostatic; antileptotic;
KW bacterial infection; DNA replication modulation.

XX Buchnera sp.

XX WO200206532-A1.

XX 24-JAN-2002.

XX 16-JUL-2001; 2001WO-US022395.

XX 14-JUL-2000; 2000US-0218246P.

XX 28-MAR-2001; 2001US-00818780.

XX (REPL-) REPLIDYNE INC.

XX Bullard JJ, Janjic N, Mchenry CS;

XX WPI; 2002-164785/21.

XX N-PSDB; ABL87941.

XX Nucleic acid encoding a DNA polymerase III holoenzyme delta subunit from
PT bacteria, useful for screening agents that modulate the subunit activity
PT which is useful in the treatment of bacterial infections e.g. S. pyogenes
PT and S. aureus.

XX Claim 19; Fig 4II; 500pp; English.

XX The present invention describes nucleic acid sequences encoding a DNA
CC polymerase III holoenzyme delta subunit (I) from bacteria. (I) has
CC antibacterial; tuberculostatic; antileptotic. Methods from the present
CC invention can be used for screening for bacterial DNA polymerase
CC holoenzyme delta subunit proteins and agents that modulate their
CC activity. The agents are useful in the treatment of bacterial infections
CC e.g. S. pyogenes, S. aureus, S. pneumoniae, Mycoplasma, Yersinia,
CC Corynebacterium, Salmonella, Mycobacterium tuberculosis or M. leprae. The
CC invention provides a convenient means of identifying compounds which
CC modulate DNA replication in bacteria and therefore provide antibacterial
CC targets, and which are also useful for amplification of DNA. ABL87935 to
CC ABL88071 and ABB84724 to ABB84816 represent sequences used in the
CC exemplification of the present invention

XX SQ Sequence 331 AA;

Alignment Scores:
Pred. No.: 0.573 Length: 331
Score: 87.00 Matches: 48
Percent Similarity: 40.18% Conservative: 40
Best Local Similarity: 21.92% Mismatches: 77
Query Match: 2.28% Indels: 54
DB: 5 Gaps: 12

US-09-931-733-1 (1-2191) x ABB84727 (1-331)

QY 621 AACCATATGAATACTTTATCAAAAAGGAGTAACTCAAGAAAAGCAAGAGGACTG 562

Db 115 AsnHisLeuSerArgPheIleGlnLysAsnLysSerLeuLysGluPheLysAsn----- 132

QY 561 AATATGGCCATAAACAATCTCTGGGTGTCGATATAATCTTTTACTGGCTTAGGGTACTC 502

```

Db      133  -----TyrAsnIleValSerCys-----PheThrProTyrAsnLeuAsn 145
QY      501  TTT---ACGTTGATACATTGAGTTGGCTTAAGCAGTAGTGTAAGAGGATATTAAAGC 445
Db      146  PheIleAsnTrpIle-----
QY      444  ACTCACTGGGAACCCCAATCCAGAGGTCAAGAGGTAGAGAACCCCTTGGAGCTTATGACCC 385
Db      151  ---LysTyrGluIleGlnGluLysLysLysLysLysLysLysLysLysLysLysLys 163
QY      384  TTCAGAACATTTTCTGCTCTTTACTT-----ATAATCATTTTATATGCA 340
Db      164  ---LysAlaPhePheLeuLeuCysLysTyrTyrGluGlyAsnThrLeuPheIleTyrLys 182
QY      339  GTAATGCTCAATGGGAGTACATACACAACTGGGTCACTATAGCGGCACTC 280
Db      183  IleuAspMetLeuPheIleIleIleIleIleIleIleIleIleIleIleIleIle 202
QY      279  AGGCAATGCTCCAT-----AATTATACCAACAAGATGAAATTTTCATA 235
Db      203  LysIleIleIleGluPhePheAspValSerProSerTyrTrpIleAsnSerIlePheGln 222
QY      234  GATCAACCAAGAAATGCTTAAATGTGC---ATTTTCCCCCACTTTGGCGAGTGGCC 178
Db      223  GlyLysThrGluLysSerPheTyrIleLeuAsnIlePhe-----PheLysLysLysTyr 240
QY      177  AACCAATGTAATTTGTCTATCATACAGAAAGTCATTAGTAGAATATACAAAATTCCT 118
Db      241  AsnProLeuLeuValArgSerLeuGlnLysAspLeuLeuGlnLeuIleHisMetLys 260
QY      117  GATAAAGCAAAATGTGCGAATATATATATATATATATATATATATATATATATATAT 58
Db      261  ArgGluLysLys-----IleSerIleTyrValMet-----LeuGluLysTyr 274
QY      57  AGTGACTAATTTATATATATATATATATATATATATATATATATATATATATATATAT 1
Db      275  Asn-----IlePheValThrArgGlyPhePheIleLysAlaPheAsnLysIle 291

```

RESULT 6
ABP39507
ID ABP39507 standard; protein; 944 AA.
XX
AC ABP39507;
XX
DT 24-JUL-2002 (first entry)
XX
DE Staphylococcus epidermidis ORF amino acid sequence SEQ ID NO:4352.
XX
KW Staphylococcus epidermidis; open reading frame; ORF; bacterial infection;
XX
OS antibacterial; gene therapy.
XX
OS Staphylococcus epidermidis.
XX
PN US6380370-B1.
XX
PD 30-APR-2002.
XX
PF 13-AUG-1998; 98US-00134001.
XX
PR 14-AUG-1997; 97US-0055779P.
XX
PR 08-NOV-1997; 97US-0064964P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX Doucette-Stamm LA, Bush D;
XX
XX WPI; 2002-381255/41.
XX
XX N-PSDB; ABN92052.
XX
XX Novel isolated nucleic acid encoding a Staphylococcus epidermidis
XX
XX PT polypeptide, useful for diagnosing and treating bacterial infections.
XX
XX PS Disclosure; SEQ ID NO 4352; 267pp; English.

```

XX      ABN90538 to ABN93374 represent Staphylococcus epidermidis open reading
CC      frame (ORF) nucleic acid sequences which encode the amino acid sequences
CC      given in ABP35124 to ABP37960. The S. epidermidis sequences have
CC      antibacterial activity and can be used in gene therapy. The sequences can
CC      also be used in the diagnosis and treatment of bacterial infections,
CC      particularly S. epidermidis infections. The sequences can be used to
CC      screen for compounds able to interfere with the S. epidermidis life cycle
CC      or inhibit S. epidermidis infection. N.B. The sequence data for this
CC      patent did not form part of the printed specification, but was obtained
CC      in electronic format directly from the USPTO web site
XX
SQ      Sequence 944 AA;

Alignment Scores:
Pred. No.: 1.13 Length: 944
Score: 86.50 Matches: 113
Percent Similarity: 34.63% Conservative: 91
Best Local Similarity: 19.19% Mismatches: 189
Query Match: 2.26% Indels: 196
DB: 5 Gaps: 26

US-09-931-733-1 (1-2191) x ABP39507 (1-944)
QY      81  TATATATGTATATATATATCGACATTTTTCCTTTATCAGGAATTTTGTATATTTCTACTA 140
Db      118  TyrLeuSerIleHisLeuMetThrTyrLeuHisTyrSerSerHisSerHisSerLeuMet 137
QY      141  ATGACT-----TTTCGTGTATGTAGC 161
Db      138  LeuThrHisPheLeuThrLeuArgLeuLeuAsnLeuLysArgLeu-CysLeuIle 157
QY      162  ACAATATACAAATGTGGCCACGTC-----CACAAAGTGGGGGAAAAATCCACA 209
Db      157  eHisLeuMetThrTyrLeuHisCysProSerArgSerHisLeuLeuMetGlnHis 177
QY      210  TTTTAAAGCATTTCTTTGCTTCATCATGATAA-----LeuCysLeuLeuIleHisLeu 241
Db      177  ysLeuMetSerMetTyrLeuLeuSerLeuLysArgLeuTyrLeuLeuSerHisSerMetM 197
QY      242  -----ATTTCATCTTTGTGG-----TATGAATTTATGGA 269
Db      197  etCysLeuTyrCysLeuThrHisSerHisSerLeuMetTrpMetHisTyrLeuMetTrpM 217
QY      270  GCATTGCTCTGAGTTGCGGCTATAGTGACCGCAGTTTGTGTATGTAGTATCCCAATT 329
Db      217  etCysLeuLeu-IleLeuAsnHis-----LeuCysLeuLeuIleHisLeu 231
QY      330  GAGACATTACTGCATATATAAATGATTATAAGTAAAGGACGAAAAAATGTTCTCAAGSGTC 389
Db      232  MetThrTyrLeuHis-----
QY      390  ATAACTCCAGGGTCTCTACCCCTTGACCTCGAATTGGGGTTCCTCAGTGAGTCCCTT 449
Db      237  -----CysProSerHisSerHis 242
QY      450  AATATCTCTTTCACCTACTGCTTAAGGCCAACTCAATGTATCCACGTAAGAGTACCC 509
Db      243  SerLeuMetLeu-MethCysLeuMetSer-MetTyrSerLeuAsnLeuLysArgSerT 262
QY      510  TAAAGCCAGTAAAGATTAT-----CACGAACACCCAGAGATCT 548
Db      262  yrLeuMetIleHisLeuTrpMetCysLeuTyrCysLeuThrHisSerHisSerLeuMetL 282
QY      549  TTATGCCCATATCCAGATCCTCTTTGCTTTTCTTGACTTACTCTCTTTTGTATAAAGT 608
Db      282  euMetHisCysLeuMetLeuMetTyrLeuLeuSerLeuLysCysLeuTyrLeuLeuSerH 302
QY      609  TATTATATGGTTTCAATCCAGATTCAGCTATTCACCTTCATGCGCCCTAGTTGATTA 668
Db      302  iserThrMetTyrLeuLeuPheProSerArgSerHisLeuLeuMetGlnMetHisTyrA 322
QY      669  AAACAATCAGCTTCTTTTGT-----CTCTGTAGATGGCTTGGACAGTACATATTAAACAGTC 725

```


Db 227 LysGluLysSerLysArgIleIleThrLeuLeuValGlnValLeuValCys----- 244
QY 528 TAATCTTTTACTGGCTTAGGTAATCTTCTAGCTGGATACATTGAGTTGGCTTAAGGCA 469
Db 245 -----PheMetProPheHisIleCysPheAlaPheLeu-----MetLeuGly 258
QY 468 GTAGTGAAGAAGGATATTAAAGGCACCTCACTGGGAACCCCAATCCAGAGGTCAAAGGTA 409
Db 259 ThrGlyGluAsnSerTyr----- 264
QY 408 GAGAACCTTGGGAAGTTATGACCCCTTCAGAACATTTTCTGCCTTTTA-----CTT 358
Db 265 ---AsnProTrpGly-----AlaPheThrThrPheLeuMetAsnLeuSerThrCysLeu 281
QY 357 ATAATCATTTTATATGAGTAATGTCCTCAATTTGGGACTAGCATACACACAACTGCGGTC 298
Db 282 AspValIleLeuTyrTrilleValSerLysGlnPheGlnAlaArgValIleSerValMet 301
QY 297 ACCTATAGGCGGCAACTCAGGACAAATGCTCCATTAATTCATACACAAAGATGAAATT 241
Db 302 LeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArgLysSerGlySerLeu 320
RESULT 8
ADC22703
ID ADC22703 standard; protein; 331 AA.
XX AC ADC22703;
XX AC
XX 18-DEC-2003 (first entry)
XX Human G protein-coupled receptor (GPCR) polypeptide #50.
XX Human; G protein-coupled receptor; GPCR; transmembrane-6 region; TM6;
XX intracellular-3 region; IC3; receptor.
XX Homo sapiens.
XX US655339-B1.
XX 29-APR-2003.
XX 13-OCT-1998; 98US-00170496.
XX 14-APR-1997; 97US-00839449.
XX 14-APR-1998; 98US-00060188.
XX 26-JUN-1998; 98US-0090783P.
XX 07-AUG-1998; 98US-0095677P.
XX (AREN-) ARENA PHARM INC.
XX Liaw CW, Behan DP, Chalmers DT;
XX WPI; 2003-742861/70.
XX N-PSDB; ADC22702.

Creating a constitutively active version of an endogenous human G protein coupled receptor (GPCR) comprises substituting a specific amino acid in the transmembrane-6 region with a different amino acid, and testing for constitutive activity.

Example 2; SEQ ID NO 184; 221pp; English.

The invention relates to a method for treating a non-endogenous, constitutively active version of an endogenous human G protein-coupled receptor (GPCR) that has a transmembrane-6 (TM6) region and an intracellular-3 (IC3) region, by substituting a specific amino acid in the TM6 region with a different amino acid, and testing for constitutive activity. The method is useful for creating a constitutively active version of an endogenous human GPCR that comprises a transmembrane 6 region and an intracellular loop 3 region. The altered human GPCR polypeptides are useful for screening test compounds for identification of inverse agonists or partial agonists of GPCR polypeptides, which may have therapeutic uses. The altered GPCRs may also be used in vivo or in

vitro in biological research. A nucleic acid encoding the altered GPCR may be used to create a transgenic animal expressing the altered GPCR. The method allows screening for compounds that modulate the activity of a human G protein-coupled receptor without the need for provision of a ligand for the receptor. This is particularly useful in allowing screening of compounds against orphan receptors for which no ligand is currently known. This sequence represents a human GPCR polypeptide of the invention.

XX SQ Sequence 331 AA;

Alignment Scores:

Pred. No.: 0.89 Length: 331
Score: 85.50 Matches: 31
Percent Similarity: 41.18% Conservative: 18
Best Local Similarity: 26.05% Mismatches: 41
Query Match: 2.24% Indels: 29
DB: 7 Gaps: 6

US-09-931-733-1 (1-2191) x ADC22703 (1-331)

QY 585 CAGAAAACCAAGAGG---ATCTGGAATATGGCCATAAACATCTCTGGGTGTCTGTA 529
Db 227 LysGluLysSerLysArgIleIleThrLeuLeuValGlnValLeuValCys----- 244
QY 528 TAATCTTTTACTGGCTTAGGTAATCTTCTAGCTGGATACATTGAGTTGGCTTAAGGCA 469
Db 245 -----PheMetProPheHisIleCysPheAlaPheLeu-----MetLeuGly 258
QY 468 GTAGTGAAGAAGGATATTAAAGGCACCTCACTGGGAACCCCAATCCAGAGGTCAAAGGTA 409
Db 259 ThrGlyGluAsnSerTyr----- 264
QY 408 GAGAACCTTGGGAAGTTATGACCCCTTCAGAACATTTTCTGCCTTTTA-----CTT 358
Db 265 ---AsnProTrpGly-----AlaPheThrThrPheLeuMetAsnLeuSerThrCysLeu 281
QY 357 ATAATCATTTTATATGAGTAATGTCCTCAATTTGGGACTAGCATACACACAACTGCGGTC 298
Db 282 AspValIleLeuTyrTrilleValSerLysGlnPheGlnAlaArgValIleSerValMet 301
QY 297 ACCTATAGGCGGCAACTCAGGACAAATGCTCCATTAATTCATACACAAAGATGAAATT 241
Db 302 LeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArgLysSerGlySerLeu 320

RESULT 9

ADA33128
ID ADA33128 standard; protein; 456 AA.

XX AC ADA33128;

XX AC
XX 20-NOV-2003 (first entry)

XX Acinetobacter baumannii protein #289.

XX Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;
XX plant biocontrol agent.

XX Acinetobacter baumannii.

XX US6562958-B1.

XX 13-MAY-2003.

XX 04-JUN-1999; 99US-00328352.

XX 09-JUN-1998; 98US-0088701P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX Breton G, Bush D;

XX WPI; 2003-576092/54.


```

XX SQ Sequence 331 AA;
Alignment Scores:
Pred. No.: 1.86 Length: 331
Score: 83.00 Matches: 29
Percent Similarity: 39.52% Conservative: 20
Best Local Similarity: 23.39% Mismatches: 47
Query Match: 2.18% Indels: 28
DB: 3 Gaps: 5

US-09-931-733-1 (1-2191) x AAY44486 (1-331)
QY 603 ATCAAAAAGGAGTACTCAAGAAAAGCAAGAGGATCTGGAATATGGCCATAACATC 544
Db 222 LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuValGlnVal 241
QY 543 TCTGGGTGTTTCGTGATAATCTTTTACTGCTTTAGGCTACTCTTACGTGGATACATTGA 484
Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255
QY 483 GTTGGCTTAAGCAGTGTGTAAGAGGATATTAAAGGCATCTCCTGGGAACCCCAATCC 424
Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264
QY 423 AGAGGTCAAGGGTAGAACCCCTTGGAGTTATGACCCCTTCAGAACATTTTCTGCCTT 364
Db 265 -----AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn 276
QY 363 TTA-----CTTATAATCATTTTATATGAGTAATGCTCAATTTGGGACTAGCATAC 313
Db 277 LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg 236
QY 312 ACACAACTCGGTACCTATAGCGGCACTCAGGAAATGTCCATAATTCATACCAC 253
Db 297 ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg 316
QY 252 AAAGATGAAATT 241
Db 317 SerGlySerLeu 320

RESULT 11
ID AAM78390
XX AAM78390 standard; protein; 331 AA.
XX AC AAM78390;
XX DT 06-NOV-2001 (first entry)
XX DE Human protein SEQ ID NO 1052.
XX XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation.
XX XX Homo sapiens.
XX OS
XX XX WO200157190-A2.
XX PN
XX XX 09-AUG-2001.
XX XX 05-FEB-2001; 2001WO-US004098.
XX XX 03-FEB-2000; 2000US-00496914.
XX PR 27-APR-2000; 2000US-00560875.
XX PR 20-JUN-2000; 2000US-00598075.
XX PR 19-JUL-2000; 2000US-00620325.
XX PR 01-SEP-2000; 2000US-00654936.
XX PR 15-SEP-2000; 2000US-00663561.
XX PR 20-OCT-2000; 2000US-00693325.
XX PR 30-NOV-2000; 2000US-00728422.
XX XX

(HYSE-) HYSEQ INC.
Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;
WPI; 2001-476283/51.
DR N-PSDB; AAK51523.
XX Nucleic acids encoding polypeptides with cytokine-like activities, useful
in diagnosis and gene therapy.
XX Claim 20; Page 3284-3285; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
encoded polypeptides (AAM78323-AAM80302) that exhibit activity relating to
cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The
polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,
e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and
activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and
inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
(AAK52582) and 3686 (AAM80020) are omitted as the relevant pages from the
sequence listing were missing at the time of publication
XX SQ Sequence 331 AA;
Alignment Scores:
Pred. No.: 1.86 Length: 331
Score: 83.00 Matches: 29
Percent Similarity: 39.52% Conservative: 20
Best Local Similarity: 23.39% Mismatches: 47
Query Match: 2.18% Indels: 28
DB: 3 Gaps: 5

US-09-931-733-1 (1-2191) x AAM78390 (1-331)
QY 603 ATCAAAAAGGAGTACTCAAGAAAAGCAAGAGGATCTGGAATATGGCCATAACATC 544
Db 222 LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuValGlnVal 241
QY 543 TCTGGGTGTTTCGTGATAATCTTTTACTGCTTTAGGCTACTCTTACGTGGATACATTGA 484
Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255
QY 483 GTTGGCTTAAGCAGTGTGTAAGAGGATATTAAAGGCATCTCCTGGGAACCCCAATCC 424
Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264
QY 423 AGAGGTCAAGGGTAGAACCCCTTGGAGTTATGACCCCTTCAGAACATTTTCTGCCTT 364
Db 265 -----AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn 276
QY 363 TTA-----CTTATAATCATTTTATATGAGTAATGCTCAATTTGGGACTAGCATAC 313
Db 277 LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg 236
QY 312 ACACAACTCGGTACCTATAGCGGCACTCAGGAAATGTCCATAATTCATACCAC 253
Db 297 ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg 316
QY 252 AAAGATGAAATT 241
Db 317 SerGlySerLeu 320

RESULT 12
AAM38949
ID AAM38949 standard; protein; 331 AA.
XX AAM38949;
XX AC AAM38949;

```

XX DT 22-OCT-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 2094.

XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX KW leukaemia.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX PF 26-DEC-2000; 2000WO-US034263.

XX PR 23-DEC-1999; 99US-00471275.

XX PR 21-JAN-2000; 2000US-00488725.

XX PR 25-APR-2000; 2000US-00552317.

XX PR 20-JUN-2000; 2000US-00598042.

XX PR 19-JUL-2000; 2000US-00620312.

XX PR 03-AUG-2000; 2000US-00653450.

XX PR 14-SEP-2000; 2000US-00662191.

XX PR 19-OCT-2000; 2000US-00693036.

XX PR 29-NOV-2000; 2000US-00727344.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;

XX PI Zhou P, Goodrich R, Drmanac RT;

XX DR WPI; 2001-442253/47.

XX DR N-PSDB; AAI58105.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders such

XX PT as central nervous system injuries.

XX PS Example 3; SEQ ID NO 2094; 10078pp; English.

XX CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the

XX CC encoded polypeptides (AAM38642-AAW42213) with nootropic,

XX CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX CC of the invention may be used to treat diseases of the peripheral nervous

XX CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX CC localised neuropathies and central nervous system diseases, such as

XX CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX CC utilisation of the activities such as: Immune system suppression,

XX CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX CC assays for receptor activity, arthritis and inflammation, leukaemias and

XX CC C.N.S disorders. Note: The sequence data for this patent did not form

XX CC part of the printed specification

XX SQ Sequence 331 AA;

Alignment Scores:

Pred. No.:	1.86	Length:	331
Score:	83.00	Matches:	29
Percent Similarity:	39.52%	Conservative:	20
Best Local Similarity:	23.39%	Mismatches:	47
Query Match:	2.18%	Indels:	28
DB:	4	Gaps:	5

US-09-931-733-1 (1-2191) x AAM38949 (1-331)

QY 603 ATCAAAAAGGTAAGTCAAGAAAAGCAAGAGGATCTGGAATATGGCCATAACATC 544

Db 222 LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuLeuValGlnVal 241

QY 543 TCTGGGTGTCGTGATAAATCTTTTACTGGCTTTAGGTACTCTTTACGTGGATACATTGA 484

Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255

QY 483 GTTGGCTTAAGCAGTAGGTGAAGAGGATATTAAGGCACCTCACTGGGAACCCCAATCC 424

Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264

QY 423 AGAGGTCAAAGGTAGAGAACCTTGGAGATTATGACCCCTTCAGAACATTTTCTGCCTT 364

Db 265 -----AsnProTrpGly-----AlaPheThrPheLeuMetAsn 276

QY 363 TTA-----CTTATAATCATTATATGAGTCAATGCTCAATTGGGACTAGCATAC 313

Db 277 LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg 296

QY 312 ACACAAACTGGGTCACTATAGCGGCAACTCAGGACAATGCTCCATAATCATACCAC 253

Db 297 ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgArgLysSerPheArg 316

QY 252 AAAGATCAAAAT 241

Db 317 SerGlySerLeu 320

RESULT 13

ABP81884

ID ABP81884 standard; protein; 331 AA.

XX AC ABP81884;

XX DT 04-MAR-2003 (first entry)

XX DE Human G protein-coupled receptor GPR18 protein SEQ ID NO:253.

XX KW G protein-coupled receptor; GPCR; antigenic peptide; gene therapy;

XX KW G protein-coupled receptor modulator; antibody; immune-related disease;

XX KW growth-related disease; cell regeneration-related disease; AIDS; cancer;

XX KW immunological-related cell proliferative disease; autoimmune disease;

XX KW Alzheimer's disease; atherosclerosis; infection; osteoarthritis; allergy;

XX KW osteoporosis; cardiomyopathy; inflammation; Crohn's disease; diabetes;

XX KW graft versus host disease; Parkinson's disease; multiple sclerosis; pain;

XX KW psoriasis; anxiety; depression; schizophrenia; dementia; memory loss;

XX KW mental retardation; epilepsy; asthma; tuberculosis; obesity; nausea;

XX KW hypertension; hypotension; renal disorder; rheumatoid arthritis; trauma;

XX KW ulcer.

XX OS Homo sapiens.

XX PN WO200261087-A2.

XX PD 08-AUG-2002.

XX PF 19-DEC-2001; 2001WO-US050107.

XX PR 19-DEC-2000; 2000US-0257144P.

XX PA (LIFE-) LIFESPAN BIOSCIENCES INC.

XX PI Burmer GC, Roush CL, Brown JP;

XX DR WPI; 2003-046718/04.

XX DR N-PSDB; ABZ42731.

XX PT New isolated antigenic peptides e.g., for G protein-coupled receptors

XX PT (GPCR), useful for diagnosing and designing drugs for treating conditions

XX PT in which GPCRs are involved, e.g. AIDS, Alzheimer's disease, cancer or

XX PT autoimmune diseases.

XX PS Disclosure; Fig 1; 523pp; English.

CC The present invention describes antigenic peptides (I) comprising: (a)
 CC any one of 1601 sequences (see ABP2019 to ABP83619) of 12-24 amino
 CC acids. Also described: (1) an assay for the detection of a particular G
 CC protein-coupled receptor (GPCR) or a candidate polypeptide in a sample;
 CC and (2) an isolated antibody having high specificity and high affinity or
 CC avidity for a particular GPCR. (I) can be used as GPCR modulators and in
 CC gene therapy. The antigenic peptides for GPCRs are useful in detecting an
 CC antibody against a particular GPCR, and in the production of specific
 CC antibodies. The peptides and antibodies are also useful for detecting the
 CC presence or absence of corresponding GPCRs. The antigenic peptides for
 CC GPCRs and antibodies are useful for diagnosing and designing drugs for
 CC treating immune-related diseases, growth-related diseases, cell
 CC regeneration-related disease, immunological-related cell proliferative
 CC diseases, or autoimmune diseases, e.g. AIDS, Alzheimer's disease,
 CC atherosclerosis, bacterial, fungal, protozoan or viral infections,
 CC osteoarthritis, osteoporosis, cancer, cardiomyopathy, chronic and acute
 CC inflammation, allergies, Crohn's disease, diabetes, graft versus host
 CC disease, Parkinson's disease, multiple sclerosis, pain, psoriasis,
 CC anxiety, depression, schizophrenia, dementia, mental retardation, memory
 CC loss, epilepsy, asthma, tuberculosis, obesity, nausea, hypertension,
 CC hypotension, renal disorders, rheumatoid arthritis, trauma, ulcers, or
 CC any other disorder in which GPCRs are involved. The antibodies may be
 CC used in immunoassays and immunodiagnosis. AB242523 to AB242869 encode
 CC GPCR proteins given in ABP81675 to ABP82018, which are used in the
 CC exemplification of the present invention

XX Sequence 331 AA;

Alignment Scores:
 Pred. No.: 1.86 Length: 331
 Score: 83.00 Matches: 29
 Percent Similarity: 39.52% Conservative: 20
 Best Local Similarity: 23.39% Mismatches: 47
 Query Match: 2.18% Indels: 28
 DB: 6 Gaps: 5

US-09-931-733-1 (1-2191) x ABP81884 (1-331)

QY	603	ATCAAAAAGGAGT	AGTCAAGAAAAGCAAGAGGATCTGGAAATATGGCCATAACATC	544
DB	222	LeuLysProLysValLysGluLysSerIleArgIleIleThrLeuLeuValGlnVal	241	
QY	543	TCGTGGGTTCGTG	GATATCTTTTACTGGCTTTAGGCTACTCTTTACGTGGATACATTGA	484
DB	242	LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu-----	255	
QY	483	GTGGGCTTAAAGG	CAGTAGTGAAGAGGATATTAAGGCACCTCACTGGGAACCCCAATCC	424
DB	256	-----MetLeuGlyThrGlyGluAsnSerTyr-----	264	
QY	423	AGAGGTCAAAAGG	TAGAACCCCTTGGAGTTATGACCCCTTCAGACATTTTCTGCCTT	364
DB	265	-----AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn	276	
QY	363	TTA-----CTTATAATCATTTTATATG	CAGTAAATGCTCAATTTGGGACTAGCATAC	313
DB	277	LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg	296	
QY	312	ACACAAACTCGGT	CACCTTACAGCGGCACTCAGGACATGCTCCATATATTCATACCAC	253
DB	297	ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg	316	
QY	252	AAAGATGAAATT	241	
DB	317	SerGlySerLeu	320	

RESULT 14
 AAB82048
 ID AAB82048 standard; protein; 333 AA.
 XX
 AC AAB82048;
 XX
 DT 22-JUN-2001 (first entry)

XX	Human G protein coupled receptor protein isomer, hGPR-iso.
DE	Human; hGPR-iso protein; dendrite cell;
XX	G protein coupled receptor protein isomer.
KW	Homo sapiens.
OS	CN1278006-A.
XX	27-DEC-2000.
XX	30-MAY-2000; 2000CN-00116229.
PF	30-MAY-2000; 2000CN-00116229.
XX	30-MAY-2000; 2000CN-00116229.
XX	(SCHR-) SOUTH CHINA RES CENT CHINA HUMAN GENE GR.
PA	Li N, Xu X, Xiao H;
PI	WPI; 2001-245686/26.
XX	N-PSDB; AAF84393.
DR	Human G protein coupled receptor protein isomer, and coding series thereof.
PT	Human G protein coupled receptor protein isomer, and coding series thereof.
XX	Claim 4; Page 12 (Disclosure); 23pp; Chinese.

CC The present sequence is human G protein coupled receptor protein isomer (hGPR-iso protein), which is expressed in human dendrite cells. The present invention also relates to a preparation method of said protein and nucleic acid sequence and a method of detecting human hGPR-iso nucleic acid sequence and polypeptide in sample

XX Sequence 333 AA;

Alignment Scores:
 Pred. No.: 1.86 Length: 333
 Score: 83.00 Matches: 29
 Percent Similarity: 39.52% Conservative: 20
 Best Local Similarity: 23.39% Mismatches: 47
 Query Match: 2.18% Indels: 28
 DB: 4 Gaps: 5

US-09-931-733-1 (1-2191) x AAB82048 (1-333)

QY	603	ATCAAAAAGGAGT	AGTCAAGAAAAGCAAGAGGATCTGGAAATATGGCCATAACATC	544
DB	224	LeuLysProLysValLysGluLysSerIleArgIleIleThrLeuLeuValGlnVal	243	
QY	543	TCGTGGGTTCGTG	GATATCTTTTACTGGCTTTAGGCTACTCTTTACGTGGATACATTGA	484
DB	244	LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu-----	257	
QY	483	GTGGGCTTAAAGG	CAGTAGTGAAGAGGATATTAAGGCACCTCACTGGGAACCCCAATCC	424
DB	258	-----MetLeuGlyThrGlyGluAsnSerTyr-----	266	
QY	423	AGAGGTCAAAAGG	TAGAACCCCTTGGAGTTATGACCCCTTCAGACATTTTCTGCCTT	364
DB	267	-----AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn	278	
QY	363	TTA-----CTTATAATCATTTTATATG	CAGTAAATGCTCAATTTGGGACTAGCATAC	313
DB	279	LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg	298	
QY	312	ACACAAACTCGGT	CACCTTACAGCGGCACTCAGGACATGCTCCATATATTCATACCAC	253
DB	299	ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg	318	
QY	252	AAAGATGAAATT	241	
DB	319	SerGlySerLeu	322	

	603	ATCAAAAAGGAGTAAGTCAAGAAAAAGCRAAAGGAGTCTCGGAATATATGGCCCATAAACATC	544
Qy	::::	::::	::::
	231	LeuIysProVallyGluSerIleAlaPheLeuValGlnVal	250
Db	::::	::::	::::
	543	TCTGGGTTCGTGATAACTTTTACTGGCTTTAGGGTACTCTTTAGCTGGATACATATGA	484
Qy		::::	
	251	LeuValCys-----PheMetPropHesileCysPheAlaPheLeu-----	264
Db		::::	
	483	GTTGGCCCTTAAGGCAGTAGGTGAAGAAGGATATTAAAGCACTACTCTGGGAACCCCAATCC	424
Qy	::::		
	265	----MetLeuGlyThrGlyGluAsnSerTyr-----	273
Db			
	423	AGAGGTCAAAGGGTAGAGAACCCTTGGAAGTTATGACCCCTTCAGAACATTTTTCTGCCCTT	364
Qy			
	274	-----AsnProTripgly-----AlaPheThrThrPheLeuMetAsn	285
Db			
	363	TTA-----CTTATAACATTTTATATGTCAGTAAATGTCCTCAATTGGGACTAGCATAC	313
Qy		::::	
	286	LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg	305
Db		::::	
	312	ACACAACCTCGGTCACTATAGCGCGCACTCAGGACAAATGCTCCATAATTCTATACCAC	253
Qy		::::	
	306	VallIserValMetLeuTyrArgAsnTyrLeuArgSerMetArgArgLysSerPheArg	325
Db		::::	
	252	AAAGATGAAATT	241
Qy			
	326	SerGlvSerLeu	329
Db			

Search completed: March 28, 2004, 08:05:09
Job time : 172 secs

RESULT 15	
AAW79374	
ID	AAW79374 standard; protein; 340 AA.
XX	
XX	
AC	AAW79374;
XX	
XX	
DT	06-NOV-2001 (first entry)
XX	
XX	
DE	Human protein SEQ ID NO 3020.
XX	
XX	
KW	Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW	vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW	tissue growth factor; immunomodulatory; cancer; leukaemia;
KW	nervous system disorder; arthritis; inflammation.

XX	Homo sapiens.
OS	
XX	
PN	WO200157190-A2.
XX	
PD	09-AUG-2001.
XX	
XX	05-FEB-2001; 2001WO-US004098.
XX	
PR	03-FEB-2000; 2000US-00496914.
PR	27-APR-2000; 2000US-00560875.
PR	20-JUN-2000; 2000US-00598075.
PR	19-JUL-2000; 2000US-00620325.
PR	01-SEP-2000; 2000US-00654936.
PR	15-SEP-2000; 2000US-00663561.
PR	20-OCT-2000; 2000US-00693325.
PR	30-NOV-2000; 2000US-00728422.
XX	(HYSE-) HYSEQ INC.
PA	
XX	
PI	Tang YT, Liu C, Drmanac RT,
PI	Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Cao Y;
PI	Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX	
XX	WPI; 2001-476283/51.
DR	N-PSDB: AAK52507.
DR	

The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication

Alignment Scores:	
Pred. NO.:	1.88
Score:	83.00
Percent Similarity:	39.52%
Best Local Similarity:	23.39%
Query Match:	2.18%
DB:	4
Length:	340
Matches:	29
Conservative:	20
Mismatches:	47
Indels:	28
Gaps:	5

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 07:59:41 ; Search time 32.5 Seconds

(without alignments)
6960.769 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09931733/runat_25032004_155918_9689/app_query.fasta_1.2375
-DB=issued_patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptp -NORM=ext -HSPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733@cgn_1_1_30@runat_25032004_155918_9689 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A.COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B.COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A.COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B.COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/pCTUS.COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	88	2.3	128	4	US-09-755-665-28
2	87	2.3	147	4	US-09-755-665-32
3	86.5	2.3	944	4	US-09-134-001C-4352
4	85.5	2.2	331	4	US-09-170-496D-184
5	85.5	2.2	456	4	US-09-328-352-4415
6	82.5	2.2	1781	2	US-08-477-451-11
7	81.5	2.1	582	4	US-09-721-870-179
8	80.5	2.1	3169	2	US-08-477-451-6
9	80	2.1	331	4	US-09-170-496D-36
10	79.5	2.1	272	4	US-09-177-419C-2
11	78	2.0	377	1	US-08-118-270-14
12	78	2.0	377	5	PCT-US93-08528-14

13	77.5	2.0	538	4	US-09-252-991A-23060	Sequence 23060, A
14	76	2.0	275	1	US-08-118-270-66	Sequence 66, Appl
15	76	2.0	275	5	PCT-US93-08528-66	Sequence 66, Appl
16	76	2.0	415	4	US-09-743-742B-6	Sequence 6, Appl
17	76	2.0	416	3	US-08-858-876A-4	Sequence 4, Appl
18	76	2.0	416	3	US-09-472-880-4	Sequence 4, Appl
19	75.5	2.0	65	4	US-09-540-236-3666	Sequence 3666, Ap
20	75.5	2.0	434	4	US-09-408-020-46	Sequence 46, Appl
21	75	2.0	284	4	US-09-543-681A-6844	Sequence 6844, Ap
22	74.5	2.0	605	2	US-08-752-307B-8	Sequence 8, Appl
23	74.5	2.0	605	4	US-09-707-802-8	Sequence 8, Appl
24	74.5	2.0	605	4	US-09-991-326-8	Sequence 8, Appl
25	74	1.9	462	3	US-09-036-987A-18	Sequence 18, Appl
26	74	1.9	462	3	US-09-370-700-18	Sequence 18, Appl
27	74	1.9	462	4	US-09-603-207-18	Sequence 18, Appl
28	73.5	1.9	1018	1	US-08-408-093-6	Sequence 6, Appl
29	73.5	1.9	1018	1	US-08-408-420A-6	Sequence 6, Appl
30	73.5	1.9	1018	1	US-08-714-901-6	Sequence 6, Appl
31	73.5	1.9	1018	3	US-08-040-741-6	Sequence 6, Appl
32	73.5	1.9	1075	4	US-09-198-452A-916	Sequence 916, App
33	73.5	1.9	3200	2	US-08-477-451-8	Sequence 8, Appl
34	73	1.9	250	4	US-09-252-991A-24614	Sequence 24614, A
35	73	1.9	434	4	US-09-408-020-14	Sequence 14, Appl
36	72	1.9	131	4	US-09-198-452A-1157	Sequence 1157, Ap
37	72	1.9	487	4	US-09-543-681A-6824	Sequence 6824, Ap
38	72	1.9	1305	4	US-08-864-785-3	Sequence 3, Appl
39	72	1.9	1353	3	US-08-894-173-2	Sequence 2, Appl
40	72	1.9	1353	3	US-09-398-193-2	Sequence 2, Appl
41	72	1.9	1353	4	US-09-473-717-3	Sequence 3, Appl
42	71.5	1.9	315	4	US-09-386-653A-9	Sequence 9, Appl
43	71.5	1.9	438	1	US-07-923-095-2	Sequence 2, Appl
44	71.5	1.9	438	1	US-08-229-511-2	Sequence 2, Appl
45	71.5	1.9	438	1	US-08-314-979-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1

US-09-755-665-28
; Sequence 26, Application US/09755665
; Patent No. 6600019
; GENERAL INFORMATION:
; APPLICANT: Pravaga, Sudhirdas K.
; APPLICANT: Majumdar, Kumud
; APPLICANT: Tailion, Bruce E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-631
; CURRENT APPLICATION NUMBER: US/09/755,665
; PRIOR FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/174,724
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-755-665-28

Alignment Scores:
Pred. No.: 0.0671
Score: 88.00
Percent Similarity: 43.62%
Best Local Similarity: 29.79%
Query Match: 2.30%
DB: 4
Length: 128
Matches: 28
Conservative: 13
Mismatches: 20
Indels: 33
Gaps: 5

US-09-931-733-1 (1-2191) x US-09-755-665-28 (1-128)

QY 230 TCATCATGAAATTCATCTTTGTGGTATGAATTATGAGCATTCCTCTGACT----- 283

```
Db 3 SerSerLeuLysIle-----TTPGlyIleLeuLeuLeuLeu 15
Qy 284 -----TCCGCGCTATAGGTGACCGACGTTGTGTGTATGCT----- 319
Db 16 CysIleLeuCysArgLeu-----CysValTyrSerAsnIleTyrTrp 30
Qy 320 -----AGTCCCAATTGACACATTACTGCATATAAATG 352
Db 31 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
Qy 353 ATTATAAGTAAAGCAGAGAAAATCTTCTGAAGGTCATACTTCCAAAGGGTCTCTACC 412
Db 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-Ile-- 69
Qy 413 CTTTGACCTCTGGATTGGGTTCCACGAGTGAGTCCTTAAT 452
Db 70 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 82

RESULT 2
US-09-755-665-32
; Sequence 32, Application US/09755665
; Patent No. 6600019
; GENERAL INFORMATION:
; APPLICANT: Prayaga, Sudhirdas K.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tallon, Bruce E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-631
; CURRENT APPLICATION NUMBER: US/09/755,665
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/174,724
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-755-665-32

Alignment Scores:
Pred. No.: 0.0966 Length: 147
Score: 87.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.27% Indels: 33
DB: 4 Gaps: 5

US-09-931-733-1 (1-2191) x US-09-755-665-32 (1-147)
Qy 230 TCATCTATGAATTCATCTTTGTGTATGATGAATATGACGATGTCTGAGT----- 283
Db 3 SerSerLeuLysIle-----TTPGlyIleLeuLeuLeuLeu 15
Qy 284 -----TCCGCGCTATAGGTGACCGACGTTGTGTGTATGCT----- 319
Db 16 CysIleLeuCysArgLeu-----CysValTyrSerAsnIleTyrTrp 30
Qy 320 -----AGTCCCAATTGACACATTACTGCATATAAATG 352
Db 31 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
Qy 353 ATTATAAGTAAAGCAGAGAAAATCTTCTGAAGGTCATACTTCCAAAGGGTCTCTACC 412
Db 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-Ile-- 69
Qy 413 CTTTGACCTCTGGATTGGGTTCCACGAGTGAGTCCTTAAT 452
Db 70 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 82
```

```
RESULT 3
US-09-134-001C-4352
; Sequence 4352, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4352
; LENGTH: 944
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4352

Alignment Scores:
Pred. No.: 0.305 Length: 944
Score: 86.50 Matches: 113
Percent Similarity: 34.63% Conservative: 91
Best Local Similarity: 19.19% Mismatches: 189
Query Match: 2.26% Indels: 196
DB: 4 Gaps: 26

US-09-931-733-1 (1-2191) x US-09-134-001C-4352 (1-944)
Qy 81 TATATATGATATATATATCGACATTTTGTCTTATCAGGAATTTTGTATATCTACTA 140
Db 118 TyrLeuSerIleHisLeuMetThrTyrLeuHisTyrSerHisSerHisLeuMet 137
Qy 141 ATGACT-----TTTCTGTATGATAGC 161
Db 138 LeuThrHisPheLeuThrLeuThrArgLeuLeuAsnLeuLysArgLeu-CysLeuLeuI 157
Qy 162 ACAAAATTACAATTGTTGGCCACTGC-----CACAAAGTGGGGGAAATTCACA 209
Db 157 eHisLeuMetThrTyrLeuHisCysProSerArgSerHisLeuLeuMetGln-MethIsc 177
Qy 210 TTTTAAAGCATTTCTTGTCTCATCTATGAA----- 241
Db 177 ysLeuMetSerMetTyrLeuLeuSerLeuLysArgLeuTyrLeuSerHisSerMetM 197
Qy 242 -----ATTTCACTTTTGTGG-----TATGAATTATGGA 269
Db 197 etCysLeuTyrCysLeuThrHisSerHisSerLeuMetTrpMetHisTyrLeuMetTrpM 217
Qy 270 GCATTGCTTGAGTTGCCGCTATAGTGACCGCAGTTGTGTGTATGTAGTCCCAATT 329
Db 217 etCysLeuLeu-IleLeuAsnHis-----LeuCysLeuLeuIleHisLeu 231
Qy 330 GAGACATTATGCAATATAAATGATTATAAGTAAAGCAGAAAATGTTCTGAGGCTC 389
Db 232 MetThrTyrLeuHis----- 236
Qy 390 ATAATTCCCAAGGGTTCTCTACCTTTGACCTCTGAGATTGGGGTCCAGTGAGTGCCTT 449
Db 237 -----CysProSerHisSerHis 242
Qy 450 AATATCCTTCTTACCTACTGCTTAAAGCCAACTCAATGATATCCAGTAAAGTACCC 509
Db 243 SerLeuMetLeu-MethHisCysLeuMetSer-MetTyrSerLeuAsnLeuLysArgSer 262
Qy 510 TAAAGCAGCAATAAAGATTAT-----CAGCAACACCCAGAGATGT 548
Db 262 yrLeuMetIleHisLeuTrpMetCysLeuTyrCysLeuThrHisSerHisSerLeuMetL 282
```


; NUMBER OF SEQ ID NOS: 8252

; SEQ ID NO 4415

; LENGTH: 456

; TYPE: PRT

; ORGANISM: Acinetobacter baumannii

US-09-328-352-4415

Alignment Scores:

Score: 0.275 Length: 456
 Pred. No.: 85.50 Matches: 63
 Percent Similarity: 39.31% Conservative: 51
 Best Local Similarity: 21.72% Mismatches: 101
 Query Match: 2.24% Indels: 75
 DB: 4 Gaps: 12

US-09-931-733-1 (1-2191) x US-09-328-352-4415 (1-456)

QY 1440 ATAGCTGAAGAGGTTCCAAACATTAGACTTTCTTACCAGCAGACCTTACATGTGAGTCA 1381
 Db 64 ileAlaGluHisAsnSerHisValGlnArg--AlaProAlaSerLeuThr---LysMet 81
 QY 1380 TTAGAGACTTATTTCCCTTAATCATGCTCAATATAAGAG---AAGCAACATGGCTTC 1324
 Db 82 MetValAlaTyrlleAlaLeuLysGluileLysAlaGlyLysLeuLysLeuAsnGluVal 101
 QY 1323 TTTCTCCCTCTTCTCTGTGACAGAGTGGCAGTGGCAGGAGGAG-----ATCTTG 1273
 Db 102 lleThrAlaThrProValValSerValValGlnThrPaspGluSerGlnMetTyrLeuLys 121
 QY 1272 GTAGGAGCATCTATTGCTACCAATCTTACTGTCATCTCTGTTGGGAAATGCTTCT 1213
 Db 122 AlaGlyGluGlnIleSerValAspGlnLeuLeuAlaGlyLeuIle---ValMetSer 139
 QY 1212 ATGAATCGCTATTGTTTAACTACTACTGAGTA-----TTCTATC 1171
 Db 140 AlaAsnAlaAlaValThrLeuAlaGluLysIleSerGlyAspValProHisPheVal 159
 QY 1170 CAATATTATAACTCAGAACGCTGGCTTCCAGATCAGAAAGTCATTATGAATATGTA 1111
 Db 160 GlnArgMetAsnGlnGluAlaGlnAlaLeuGlyMetLysAspThrHisPheSer----- 177
 QY 1110 AATCTTCTTAAATTTATAGAGAGCTTTTAACACTAGCTGTCTCTTGACCATGACTA 1051
 Db 178 AsnProAla-GlylleThrMetProAspHisTyrThrThraIa-----HisAspLe 194
 QY 1050 GGTACTTACTCTCAATGATTAGTATCAGATGCTCAGATTTCTGCATTTATCTCTGGA 991
 Db 194 uSerLeuLeuSerGlnAlaVal-IleHisGlnThrProGluTyrLeuHisTyrSerLysM 214
 QY 990 TA----- 989
 Db 214 etProSerPheSerTyrAsnGlnArgPheHisAlaThrAsnLeuAlaLeuLysTyrA 234
 QY 988 -----ACAAACGTGGATGATGAGTGGACTCTCAACA 961
 Db 234 spProSerValAspGlyLeuLysThrGlyTyrThrLysAlaAlaGlyTyrAsnLeuAla 254
 QY 960 TCACAGCATCGAAGACTTGAATACTCGTGAATAATGCACACTCCCGTAGTGTGAGGA 901
 Db 254 euThrAlaSer-----Argp 259
 QY 900 GTCTGTTTAAATCCTCTGTCAGAGTGTATTTTCTCTGTTAAGACACTGGGTCCTTC 841
 Db 259 roSerPheSerProAsnLeuProGlnArgArgLeuLeuValIleValLeuGlyThrProS 279
 QY 840 TAGCGTGTCTTAACCCAGGATGTCAATTAATGTCATCAACTAGCA----- 794
 Db 279 erAlaValLysArgAlaGluIleAlaAspLysLeuMetAsnLeuAlaTyrAlaTyrThra 299
 QY 793 --GATGATGTGGTGTAAACCCAGCAGCAGATCTGCAACGAGGATGTGTCATGAGTGA 736
 Db 299 rgAspGluValValIleProGluGlnLysLeuIleAlaGluLeuProValIleLysSer- 318

QY 735 AGGCCCCATAGACTGTTTAATATGTAC 710
 Db 319 -----ThrLeuLysMetPhe 323

RESULT 6

US-08-477-451-11
 ; Sequence 11, Application US/08477451
 ; Patent No. 5928865
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
 ; NUMBER OF SEQUENCES: 46
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,451
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0335.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 510-601-2708
 ; TELEFAX: 510-655-3542
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1781 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-477-451-11

Alignment Scores:
 Pred. No.: 1.36 Length: 1781
 Score: 82.50 Matches: 75
 Percent Similarity: 32.97% Conservative: 45
 Best Local Similarity: 20.60% Mismatches: 109
 Query Match: 2.16% Indels: 135
 DB: 2 Gaps: 18

US-09-931-733-1 (1-2191) x US-08-477-451-11 (1-1781)

QY 243 TTTTCATCTTTGTGTATGAATTATGAGCATTGTCCTGAGTTGGCCTATAGTGCACCG 302
 Db 1041 PheHisLeuLysCys-----PheGlyAsnProArgPheLysAlaValPheArgPheGlu 1057
 QY 303 CAGTTTGTGTGTATG-----CTAGTCCCAATTGAGACATTACTG 341
 Db 1058 SerValLeuLeuValHisPheGluLeuAspHisPheValIlePheLeuGluArgAspLeu 1077
 QY 342 CATATAAATGATTATTAAGTAAAGGCAGAAATAATGTTCTGAAGGTCATAACTTCAAG 401
 Db 1078 HisVal----- 1079
 QY 402 GGTTCCTCTACCTTTGACCTCTGGAATTGG-----GTT 434
 Db 1080 -----LeuAspPheCysLeuGlyIleAsnPheLeuPhePheHisPheAspPheAspIle 1097
 QY 435 CCCAGTGAAGTCCTTAATATCTCTTCTTCCACTACTGCTTAAAGCCCAACTCAATGTATCC 494
 Db 1098 ProGlnCysThrLysArgArgLeuAsnThrLeuPro--ArgGlnAlaArgCysLeuA 1117

QY	495	ACGTAAAGAGTACCCTAAAGCCAGTAAAA	----	GATTATCACGAACACCCAGAGATGTTTA	551
Db					
Db	1117	snLeuGlnLysThr	-----	ProLeuSerCysHisHisHisArgHisProLeuLeuLeuThrA	1135
QY	552	TGGCCATATTCCAGATCCTCTTTTGCTTTTCTTGACTTACTCCTTTT	TTTGATAAAGTTAT	611	
Db					
Db	1135	rgSer	-----		1136
QY	612	TCATATGTTTCAAATCCAGAAATTCAGCTATTCACTTCATGGCCCTAGTTTGGATTAAAA	671		
Db					
Db	1137	TrpLeuGlnAlaGluSerSerIleLeuProLeuPheAlaLeuGluTrpLeu	----	1153	
QY	672	CAATCAGCTTTCTTTTCTCTGTAGATGGCCTTGGACAGTACATATTAAACAGTCTATGGG	731		
Db					
Db	1154	PheLeu	-----	Trp	1163
QY	732	CCCTTTCCACTCATGACCAC	-----	ATCCTGCTTGCAGATCTCGTCTGG	776
Db					
Db	1163	euleuIleArgLeuGluArgPhePheValTrpLeuValArgLeuProGluTrpLeuLeu	1182		
QY	777	GTTAACACCACATCATCTGCTAGTGTGATGCCATTTATTTCACATCTCTGGGGTTACAGCAC	836		
Db					
Db	1183	ArgGlnValGlnAlaLeuLeuProLeuTyr	-----	1193	
QY	837	GCTAAAGGCACCCAGTGTCTTAAACAGAGAAAATACACTGCTGACAGAGAGTTAAAC	896		
Db					
Db	1194	HisSerGluThrAsnHisLysLysSerArgLysLysIleCysLeuLysA	1210		
QY	897	AGACTCCTCACACTACCGGAAGTTCAGTGCATTTCCAGGATTTTCCAGTCTTCGATGCT	956		
Db					
Db	1210	snIleSerHisLeu	-----	PheProHisPheThrIleIleLeuValC	1224
QY	957	GTCATGTTGAG	-----	AGTCCCATATCCACGTTTGTATTATCCAGGAATAA	1001
Db					
Db	1224	yethrLeuSerValLysLeuSerPheCysProLeuSer	-----	1236	
QY	1002	TGCAGAAATCTCAGCATTCGTGAATCTAATCATTCAGGTAAAGTAAACCTAGTCATGGTC	1061		
Db					
Db	1237	LeuSerLeuMetSerArgValLys	-----	1244	
QY	1062	AAGAGCACAGCTAGTGTTTAAACTCTTCATATAAATTTAAGGAAGGATTTACATATTCAAT	1121		
Db					
Db	1245	CysAspPheLysLysAspLeuPheAlaThrIleLeuLeuTyrSerG	1260		
QY	1122	AATGACTTCTGATCTG	-----	GAAGCCCGAGCTTGTTCAGTTATTA	1163
Db					
Db	1260	luLeuMetLeuGluIleAlaValSerValLeuValAlaAsnProAspLeuSerIleIleG	1280		
QY	1164	ATAATTGGATGAATACCTCAGGTAGTATTTAAACAATAGCGATTCATAGAGACATTT	1223		
Db					
Db	1280	luAlaPheCysAsnLysLeuValLeuPhePheThrAsnSerIleTyrSerLeuLysIleS	1300		
QY	1224	CCCCAAC	1231		
Db					
Db	1300	erAlaThr	1302		

RESULT 7
US-09-721-870-179
; Sequence 179, Application US/09721870
; Patent No. 6632621
; GENERAL INFORMATION:
; APPLICANT: Lowery, David E.
; APPLICANT: Geary, Timothy G.
; APPLICANT: Kubiak, Teresa M.
; APPLICANT: Larsen, Martha J.
; TITLE OF INVENTION: MODULATORS OF G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 28341/6223
; CURRENT APPLICATION NUMBER: US/09/721.870
; CURRENT FILING DATE: 2000-11-24
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 179

```

; LENGTH: 582
; TYPE: PRT
; ORGANISM: C. elegans
US-09-721-870-179

Alignment Scores:
Pred. No.:      0.996      Length:      582
Score:          81.50      Matches:     68
Percent Similarity: 33.43%      Mismatches: 42
Best Local Similarity: 20.67%      Indels:     98
Query Match:     2.13%      Gaps:      121
DB:              4          Gaps:      15

US-09-931-733-1 (1-2191) x US-09-721-870-179 (1-582)

QY 1306 CAGGAAAGGAGGGAGAGAAGCAGCATGTCCTCTCTCTATATTATTGAGCATCATTTAAGGG 1365
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 GlnThrArgSerHisAlaTyAlaMetIleThrLeuValTrpValIleAlaIleAla 149

QY 1366 AATATAAGTCTCTAATGACTC-----ACATGTAAGTCTGCTGGTAAAGAAGT 1413
      ||||| ||| :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 150 AsnIleLeuMetLeuPheMetTyrGluGlnGlnThrTyr-SerSerAsnGlyTyrThrCys 169

QY 1414 TCTAATGCTTGGACTCCT-----TTTCAGTATCGAATTTTGTAGAAAAAATTCTTA 1464
      ||| :: :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 170 AlaProIleHisProIleTyrHisPheAlaTyrGln-----GluAsnAglYs 186

QY 1465 TTTGTGTTTATATACATTTATGAGATACAAAGTGTAATTCGTTTACATGGATATATCCAT 1524
      |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 Phe-----His 188

QY 1525 ATTGGTGAAGTCAGAGATTTCACTGTGCACATCACCCGAAAAATGTTAACTGACCCATT 1584
      :: :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 LeuLysSerIleIleTyrPheGlnValTyrMetThrValValLeuLeuValIleProLeu 208

QY 1585 AAGATA-----1590
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 209 ValValMetAlaGlyLeuTyrGlyAsnValIleThrSerLeuLysSerGlyIleLysLeu 228

QY 1591 ---ATTTCATCCCCCATTTCCCTCCACACATGCTCCAGCCCTTGCCAGGAAACTGTT 1647
      ||| :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 GluIleAlaSerValAspProProLeuAlaThrAla-----240

QY 1648 CATTTCTCTGACTAACAGAAAGCAAGCACTAAACACACTGGTGGAGG-----AGTCCTC 1701
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 -----ThrThrThrGlyAlaLysAsnLeuGlySerHisSerAspSerAla 255

QY 1702 CACATTGTTCTACTCCATTTCTCTGGGAATAGCAGATAGGACGACGACGACTA 1761
      ||| :: :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 HisLeuLeuLeuAsnValLeuValGlySerSerGln-----268

QY 1762 GTCAGCTAACTGACGACCAACCAAGGCCTTTTTCCTTGTTATCTTTGCAGATACTTC 1821
      ||||| :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 -----LeuAlaArgAlaThrSer-----CysIleAlaLeuAsnThrPhe 281

QY 1822 ATTTCTTAGCGTTTCTGGAGATTACAATCTCGGTTCCGTTTCTGGGAACCTTTACT 1881
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 282 -----SerCysAsnAsnThrThrAsnProPheSer 291

QY 1882 GATTTATCTCCCCCTCACACAAATAAGCATTGATTCTCTGCATTTCTCAAGATCTCAAGA 1941
      ||| :: :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 292 ThrLeuThrProPro-----ProIleGlnGlnLysAsnArg 304

QY 1942 TCTGCACTACTGTTGAAAAAATTTCCAGGTGAG-----1974
      ||| ||| :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 305 SerLysProGlnLeuLeuGlnLeuProGlyLysIleAspAsnPheGluGluPheArgLysLeu 324

QY 1975 TACTGTTCTCGATTTGTAAATATGAT-----CTTGTCTCT 2010
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 325 GlnCysLeuSerAspCysArgSerAspGlyValLeuPheProProAlaIleValAla 344

QY 2011 TCCTTGAAGTCCGAG-----AATCAAGGGGACATC 2043
      ||| ||| :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 345 SerMetThrAspGluGlnLysLeuSerPheTrpAsnLysLeuSerAsnLysLeuThrPhe 364
QY 2044 AGTATTGTTATTACAGGTTCATGGATGATGGAGTAGGCTGAGTATTACAGAAAGTGAA 2103
Db 365 Ser-----GlnGlnAspLysThrValGlnHisProAsnPheGlyHisArgLysSerAsp 382
QY 2104 ACTGAGTTCCTTCATATGATGATCCTCA 2130
Db 383 ThrSerIleCysLeuGluAsnProSer 391
RESULT 8
US-08-477-451-6
; Sequence 6, Application US/06477451
; Patent No. 5928865
; GENERAL INFORMATION:
; APPLICANT: Covacci, Antonello
; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: CA
; COUNTRY: USA
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,451
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McClung, Barbara G.
; REGISTRATION NUMBER: 33,113
; REFERENCE/DOCKET NUMBER: 0335,002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 510-601-2708
; TELEFAX: 510-635-3542
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3169 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-477-451-6
Alignment Scores:
Pred. No.: 3 32 Length: 3169
Score: 80.50 Matches: 44
Percent Similarity: 30.39% Conservative: 18
Best Local Similarity: 21.57% Mismatches: 51
Query Match: 2.10% Indels: 91
DB: 2 Gaps: 9
US-09-931-733-1 (1-2191) x US-08-477-451-6 (1-3169)
QY 530 CACGAACACCCAGAGATGTTATGGCCATATCCAGATCCTCTGCTTT---TCTTGA 586
Db 1344 HisHisCysProHisThrThrMetLysLeuPheGlnValLeuCysHisTyrAspSerPro 1363
QY 587 CTTACT-----CCTTTTGTATAAGTATTATCATGTTTCAA 625
Db 1364 LeuThrGlnLysThrGluArgTyrCysGluSerIleIleAspThrCysAsnTrpAspGln 1383
QY 626 ATCCAGAAATTCAGTATTAC----- 646
Db 1384 CysGlnArgLeuProGlnHisLeuThrPheProTyrArgGlnGluLeuLysAsnTyrPro 1403
QY 647 TTCATGGCCCTAGTTTGGATTAAACAAATCAGCTTTCTTTTGTCTGTAGATGGCCTTGG 706

Db 1404 PheHisGlnValAsnLeuLeuGlnAlaValSer-----LeuValCysArgTrpTrpTrp 1421
QY 707 ACAGTACATATTAAACAGTCTATGGGCCCTTTCCATCATGACCACATCTGCTTTCAGAT 766
Db 1422 LeuTyr-----LeuValGlyHisGlyGluLeuValLeu--Aap 1433
QY 767 CTGCTGCTGGGTTAAACACCATCATCTGCTAGTGTGATGTCATTTATTGACATCCTGGG 826
Db 1434 LeuLeuLeuHisGlnLysIleValPheCysGlnIleMet----- 1446
QY 827 GTTACAGCAGCTAAAGGGCACCCAGTGTCTTAACAGAGAAAAATACACTGCTGAGCAG 886
Db 1447 -----ArgGluLysTyrProThrAsnPro 1454
QY 886 ----- 886
Db 1455 HisGlnAlaLeuHisPhePheLeuLeuLeuAlaPheLeuSerArgThrHisGlnThrArg 1474
QY 887 -----AGGATT 892
Db 1475 ProThrPheAspProPheGlyPheAspSerLeuPheLeuGlnLysIleValTrpGlnIle 1494
QY 893 AAACAGACTCCT-----CACACTACCGGAAGTTCAGTGCATTTCCAGGATTTTCA 943
Db 1495 SerLysThrProLeuSerLeuTyrHisGlnGlyGlyValValTyrPheProAsnPheGly 1514
QY 944 AGTCTTCGATGC 955
Db 1515 SerIleGlyCys 1518
RESULT 9
US-09-170-496D-36
; Sequence 36, Application US/09170496D
; Patent No. 6555339
; GENERAL INFORMATION:
; APPLICANT: Behan, Dominic P.
; APPLICANT: Chalmers, Derek T.
; APPLICANT: Liaw, Chen W.
; TITLE OF INVENTION: No. 6555339-Endogenous, Constitutively Activated Human G Protein
; FILE REFERENCE: AREN-0040
; CURRENT APPLICATION NUMBER: US/09/170,496D
; CURRENT FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 294
; SOFTWARE: Patent In version 3.1
; SEQ ID NO 36
; LENGTH: 331
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-170-496D-36
Alignment Scores:
Pred. No.: 1.13 Length: 331
Score: 80.00 Matches: 28
Percent Similarity: 39.52% Conservative: 21
Best Local Similarity: 22.58% Mismatches: 47
Query Match: 2.10% Indels: 28
DB: 4 Gaps: 5
US-09-931-733-1 (1-2191) x US-09-170-496D-36 (1-331)
QY 603 ATCAAAAAGGAGTAGTCAAGAAAAAGCAAGAGAGATCTGGAATATGCCATAAACATC 544
Db 222 LeuLysProLysValLysGluLysSerIleArgIleIleIleThrLeuLeuValGlnVal 241
QY 543 TCTGGGTGTCGTGATAATCTTTTACTGGCTTAGGGTACTCTTTACGTGGATACATGCA 484
Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255
QY 483 GTTGGCCTTAAGCGAGTAGGTGAAGAGATATTAAAGCACTACTGGAACCCCAATCC 424
Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264


```

Db      240 AlaGlyLysArgArgProSerArgLeuValAla-----LeuArgGlu 253
QY      1368 TATAAGTCTCTAATGACTCACATGTAAGTCTGGTGAAGAAAGTCTTAATGTTGGAAC 1427
Db      254 GlnLysAlaLeuLysThrLeuGlyIleMetGlyValPheThrLeu---CysTrp-Le 272
QY      1428 TCCTTTACGCTATCGAATTTGTAAAGAAAAAACTTATTGTTTATATACATTATGA 1487
Db      272 uproPhePheHisArgGluLeuValProAspArgLeuPheValPhe-PheAsnTrpLeuA 292
QY      1488 GATACAAGTGAATTCGTTACATGATATATGCGATATGCGTGAAGTCAAGATTTTCAG 1547
Db      292 rg-----TyrAlaAsnSerAlaPheAsnProIleIleT 303
QY      1548 TGTGCACATCACCCGAA 1564
Db      303 YrCysArgSerProAsp 308

RESULT 12
; Sequence 14, Application PC/TUS9308528
; GENERAL INFORMATION:
; APPLICANT: New York University
; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
; NUMBER OF SEQUENCES: 348
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08528
; FILING DATE: 09-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/943,236
; FILING DATE: 10-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, Kevin G.
; REGISTRATION NUMBER: 34,033
; REFERENCE/POCKET NUMBER: MURPHY=2 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 377 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US93-08528-14

Alignment Scores:
Pred. No.: 2,17 Length: 377
Score: 78.00 Matches: 35
Percent Similarity: 43.93% Conservative: 12
Best Local Similarity: 32.71% Mismatches: 34
Query Match: 2.04% Indels: 26
DB: 5 Gaps: 5

US-09-931-733-1 (1-2191) x PCT-US93-08528-14 (1-377)
QY      1272 CCAAGATCTCTCTCT-----GCCACTGCCACTCTGTGCACAGGAAG 1313

```

```

Db      220 ProGlyProProArgProAlaAlaAlaAlaThrAlaProLeuAlaAsnG.Yarg 239
QY      1314 GAGGGAGAAGA-----AGCCATGTTGCTTCTCTTATTTGAGCATCATTTTAAGGGA 1367
Db      240 AlaGlyLysArgArgProSerArgLeuValAla-----LeuArgGlu 253
QY      1368 TATAAGTCTCTAATGACTCACATGTAAGTCTGGTGAAGAAAGTCTTAATGTTGGAAC 1427
Db      254 GlnLysAlaLeuLysThrLeuGlyIleMetGlyValPheThrLeu---CysTrp-Le 272
QY      1428 TCCTTTACGCTATCGAATTTGTAAAGAAAAAACTTATTGTTTATATACATTATGA 1487
Db      272 uproPhePheHisArgGluLeuValProAspArgLeuPheValPhe-PheAsnTrpLeuA 292
QY      1488 GATACAAGTGAATTCGTTACATGATATATGCGATATGCGTGAAGTCAAGATTTTCAG 1547
Db      292 rg-----TyrAlaAsnSerAlaPheAsnProIleIleT 303
QY      1548 TGTGCACATCACCCGAA 1564
Db      303 YrCysArgSerProAsp 308

RESULT 13
; Sequence 23060, Application US/09252991A
; Patent No. 5551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23060
; LENGTH: 538
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23060

Alignment Scores:
Pred. No.: 3,03 Length: 538
Score: 77.50 Matches: 57
Percent Similarity: 36.33% Conservative: 44
Best Local Similarity: 20.50% Mismatches: 89
Query Match: 2.03% Indels: 89
DB: 4 Gaps: 11

US-09-931-733-1 (1-2191) x US-09-252-991A-23060 (1-538)
QY      104 ATTTTCTCTTATCAGGAATTTTGTATATTTCTACTAATGACTTTTCTGTATGATACAC 163
Db      129 lPheAlaPheGlyGlyCysAlaLeuPheAlaThrSerTyrTyrValVal----- 145
QY      164 AAATTACAA-----TTGTTGGCCACTGCACAAAG 193
Db      146 -----GlnArgThrSerGlnAlaArgLeuLeuSerAspThrLeuAlaAlaPheThrPhe 163
QY      194 TGGGGGAAAAATGCACATT-----TTTAAGCATTTCTTTGTTCTCA 232
Db      164 TrpGlyTrpGlnAlaValIleValGlyAlaValLeuThrLeuProGlnGlyPheThrThr 183
QY      233 TCTATGAAATTTTCATCTTTGTGTATGAATTTGAGCATTTGCTCTGAGTTCGCCCTA 292
Db      184 SerLysGluTyrAlaGluLeuGluTrpProLeu-----AlaIleLeuLeuAlaIleVal 201
QY      293 TAGGTGACCCGACGTTTGTGTATGTGTAGTCCCAATTGACACATTACTGCATATATAAATG 352

```

Db 202 TrpIleThrTyAlaIleValPhe-----PheGlyThr 212
 QY 353 ATTAAAGTAAAGGACAGAAATCTCTGAAGGCTCATAACTTCCAAAGGTT----- 405
 Db 213 IleValLysArgLysValHisIleTyValGlyAsnTrpPheTyGlyAlaPheIle 232
 QY 406 -----CTCTACCTTTGACCTCTGCTGATGGGTTCCCAAGTGGAGTGCCTTAATA 453
 Db 233 LeuValThrAlaMetLeuHisIleValAsnHisMetSerLeuProValSerTrpPheLys 252
 QY 454 TCCTTCTTCACTACTGCTTAAAGGCCAACCTCAATGATTCACGTAAAGAGTACCTTAA 513
 Db 253 SerTySerAlaTySerGlyAlaThrAspAlaMetVal----- 265
 QY 514 GCCAGTAAAGATTATACAGAACACCCAGAGATGTTATGGCCATATTCCAGATCCTCTT 573
 Db 266 -----GlnTrpTrpTyGlyHis-----AsnAlaVal 274
 QY 574 TGCCTTTCTTACTCTCTCTTTTGCATAAGTTATTCATATGG----- 619
 Db 275 GlyPhePheLeuThr-ThrGlyPheLeuGlyMetMetTyTrpPheValProLysGlnAl 294
 QY 620 -----TTTCAAAATCCAGAAATTCAGCTATTCACTTATGCGCCCTAGTT----- 661
 Db 294 aGluArgProValTySerTyArgLeuSerIleValHisPheTrpAlaLeuIleSerLe 314
 QY 662 -----TGGATTAAGAAC 672
 Db 314 uTyIleTrpAlaGlyProHisHisLeuHisTyThrAlaLeuProAspTrpAlaGlnSe 334
 QY 673 AATCAGCTTT-----CTTTTGCTCTGTAGATGGCTTGGACAGTACATATAACAG 723
 Db 334 rLeuGlyMetValMetSerLeuIleLeuAlaProSerTrpGlyGlyMetIleAsnGl 354
 QY 724 TCTATGGCCCTTTCCACTCATGACCATCTCTGTCGAGATCTGCTGCTG 775
 Db 354 yMetMetThrLeuSerGlyAlaTrpHisLysLeuArgThrAspProIleLeu 371

RESULT 14

US-08-118-270-66
 ; Sequence 66, Application US/08118270
 ; Patent No. 5508384
 ; GENERAL INFORMATION:
 ; APPLICANT: Murphy, Randall B.
 ; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
 ; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
 ; NUMBER OF SEQUENCES: 348
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/118,270
 ; FILING DATE: 09-SEP-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/943,236
 ; FILING DATE: 10-SEP-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Townsend, Kevin G.
 ; REGISTRATION NUMBER: 34,033
 ; REFERENCE/DOCKET NUMBER: MURPHY-2A
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528

TELEX: 248633
 ; INFORMATION FOR SEQ ID NO: 66:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 275 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; US-08-118-270-66
 Alignment Scores:
 Pred. No.: 3.26 Length: 275
 Score: 76.00 Matches: 37
 Percent Similarity: 43.17% Conservative: 23
 Best Local Similarity: 26.62% Mismatches: 46
 Query Match: 1.99% Indels: 33
 DB: 1 Gaps: 7
 US-09-931-733-1 (1-2191) x US-08-118-270-66 (1-275)
 QY 524 GATTATCAGCAACACCCAGAGATGTTATGGCCATATTCCAGATCCTCTTTGCT----- 577
 Db 27 AspSerHisLeuHisThrProMetTyLeuPheLeuSerAsnLeuSerPheSerAspLeu 46
 QY 578 ---TTTCTTTGACTTACTCTCTTTTTCATAAAGTTATTCATATGTTTCAATCCAGAT 634
 Db 47 CysPheSerSerValThr-----MetLeuLysLeuGlnAsnIleGlnSerGlnVal 64
 QY 635 TCAGCTATTCACTTCATGCGCCCTAGTTTGGATTAAACAAATCAGCTTTCTTTGCTCTGT 694
 Db 65 ProSerIleSerTyAlaGlyCysLeuTip-----IlePhePheLeuLeuPhe 81
 QY 695 AGATGGCCCTTGGACAGTACATATTAACAGTCTATGG-----CC 733
 Db 82 GlyTyLeuGlyAsnPheLeuLeuValAlaMetAlaTyAspArgTyValAlaIleCy 101
 QY 734 CTTTCCACTCATGACACACA-----TCCTGCTTGCAGATCTG 769
 Db 101 sPheProLeuHisTyThrAsnIleMetSerHisLysLeuCysThrCysLeuLeuVal 121
 QY 770 CTGCTGGGTTAAACACACATCATCTGCTAGTTGATGCACTTTATTGACATCTCTGGGTT 829
 Db 121 lPheTrpIleMetArgSerSerHisAlaMet-----MetIleThrLe 135
 QY 830 ACAGCAGCTTAAAGGCCACCCAGTGTCTTAAACAGAGAAAAATACACTGCTGAC 884
 Db 135 ulleAlaAlaArg-----LeuSerPheCysGluAsnValLeuLeuAsn 150

RESULT 15

PCT-US93-08528-66
 ; Sequence 66, Application PC/TUS9308528
 ; GENERAL INFORMATION:
 ; APPLICANT: New York University
 ; TITLE OF INVENTION: POLYPEPTIDES OF G-COUPLED PROTEIN
 ; TITLE OF INVENTION: RECEPTORS, AND COMPOSITIONS AND METHODS THEREOF
 ; NUMBER OF SEQUENCES: 348
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/08528
 ; FILING DATE: 09-SEP-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/943,236

Search completed: March 28, 2004, 08:15:04
Job time : 50.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 08:12:12 ; Search time 140.5 Seconds
(without alignment)
8160.868 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1065169 seqs, 261661801 residues

Total number of hits satisfying chosen parameters: 2130338

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p model -DEV=xlib
-O=/cgn2_1/USPO.spool/US09931733/runat_25032004_155919_9779/app_query.fasta_1.2375
-DB=published Applications AA_QPWT=fasten -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09931733 @CIGN 1.1.155 @runat_25032004_155919_9779
-NCPU=6 -ICPU=3 -NO MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA:

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pcp.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pcp.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pcp.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pcp.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pcp.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pcp.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pcp.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pcp.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pcp.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pcp.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pcp.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pcp.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pcp.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pcp.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pcp.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pcp.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pcp.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pcp.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

1	88	2.3	128	9	US-09-755-665-28	Sequence 28, Appl
2	88	2.3	128	12	US-10-445-641-2	Sequence 2, Appl
3	87	2.3	147	9	US-09-755-665-32	Sequence 32, Appl
4	87	2.3	147	12	US-10-445-641-6	Sequence 6, Appl
5	87	2.3	331	11	US-09-906-179A-12	Sequence 12, Appl
6	85.5	2.2	331	14	US-10-251-385-184	Sequence 184, App
7	85.5	2.2	357	12	US-10-424-599-267968	Sequence 267968,
8	83	2.2	331	9	US-09-826-508-16	Sequence 16, Appl
9	83	2.2	331	14	US-10-225-567A-253	Sequence 4, Appl
10	83	2.2	331	14	US-10-225-567A-253	Sequence 253, App
11	82.5	2.2	310	9	US-09-771-209-75	Sequence 75, Appl
12	81	2.1	34	9	US-09-764-860-328	Sequence 328, App
13	81	2.1	34	14	US-10-074-095-328	Sequence 328, App
14	81	2.1	34	15	US-10-212-872-328	Sequence 328, App
15	80	2.1	312	9	US-09-771-209-77	Sequence 77, Appl
16	80	2.1	331	14	US-10-251-385-36	Sequence 36, Appl
17	80	2.1	331	15	US-10-411-010-18	Sequence 18, Appl
18	80	2.1	355	12	US-10-425-114-54458	Sequence 54458, A
19	79	2.1	362	14	US-10-017-161-2364	Sequence 2364, Ap
20	79	2.1	362	15	US-10-292-798-2008	Sequence 2008, Ap
21	78.5	2.1	65	12	US-10-424-599-23444	Sequence 23444,
22	78.5	2.1	473	15	US-10-369-493-2432	Sequence 2432, Ap
23	78	2.0	122	12	US-10-424-599-150480	Sequence 150480,
24	78	2.0	252	9	US-09-974-298-193	Sequence 193, App
25	78	2.0	457	12	US-10-425-114-54364	Sequence 54364, A
26	78	2.0	490	12	US-10-425-114-43063	Sequence 43063, A
27	78	2.0	646	12	US-10-425-114-65028	Sequence 65028, A
28	77.5	2.0	475	12	US-10-282-122A-66300	Sequence 66300, A
29	77	2.0	356	12	US-10-425-114-67444	Sequence 67444, A
30	77	2.0	614	12	US-10-425-114-71790	Sequence 71790, A
31	77	2.0	1407	15	US-10-369-493-1528	Sequence 1528, Ap
32	76.5	2.0	277	14	US-10-017-161-2342	Sequence 2342, Ap
33	76.5	2.0	658	14	US-10-156-761-14580	Sequence 14580, A
34	76	2.0	263	14	US-10-017-161-2420	Sequence 2420, Ap
35	76	2.0	263	15	US-10-292-798-2060	Sequence 2060, Ap
36	76	2.0	289	12	US-10-425-114-41818	Sequence 41818, A
37	76	2.0	416	14	US-10-205-219-21	Sequence 21, Appl
38	76	2.0	437	14	US-10-017-161-2204	Sequence 2204, Ap
39	76	2.0	437	15	US-10-292-798-1850	Sequence 1850, Ap
40	76	2.0	446	12	US-10-424-599-215419	Sequence 215419,
41	76	2.0	859	10	US-09-978-522-3	Sequence 3, Appl
42	76	2.0	862	10	US-09-978-522-1	Sequence 1, Appl
43	75.5	2.0	226	14	US-10-017-161-2368	Sequence 2368, Ap
44	75.5	2.0	342	15	US-10-369-493-12383	Sequence 12383, A
45	75.5	2.0	434	13	US-10-027-806-46	Sequence 46, Appl

ALIGNMENTS

RESULT 1
US-09-755-665-28
; Sequence 28, Application US/09755665
; Patent No. US20020107186A1
; GENERAL INFORMATION:
; APPLICANT: Prayaga, Sudhirdas K.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tailon, Bruce E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-631
; CURRENT APPLICATION NUMBER: US/09/755,665
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/174,724
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 28
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-755-665-28

```
Alignment Scores:
Pred. No.: 0.671 Length: 128
Score: 88.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.30% Indels: 33
DB: 9 Gaps: 5

US-09-931-733-1 (1-2191) x US-09-755-665-28 (1-128)
QY 230 TCATCTATGAAATTTTCATCTTTGGTATGAATATGAGCATTGCTCGAGT----- 283
DB 3 SerSerLeuLysLeu-----TrpGlyLeuLeuAlaLeuLeu 15
QY 284 -----TGCAGCTATAGTGACCGCAGTTTGTGTGTATGCT----- 319
DB 16 CysLeuLeuCysArgLeu-----CysValTyrSerAsnAsnIleTyrTrp 30
QY 320 -----AGTCCCAATTGAGACATTACTGCATATAAAATG 352
DB 31 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
QY 353 ATTATAAGTAAAGGACAGAAAATGTTCTGAAGGGTCATAACTTCCAAGGGTCTCTTACC 412
DB 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-Ile-- 69
QY 413 CTTTGACCTCTGGATTGGGGTCCAGTGAGTGCCTTAAT 452
DB 70 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 82

RESULT 3
US-09-755-665-32
; Sequence 32, Application US/09755665
; Patent No. US20020107186A1
; GENERAL INFORMATION:
; APPLICANT: Prayaga, Sudhirdas K.
; APPLICANT: Majumder, Kumud
; APPLICANT: Tailon, Bruce E.
; APPLICANT: Spaderna, Steven K.
; APPLICANT: Spyttek, Kimberly A.
; APPLICANT: MacDougall, John
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-631
; CURRENT APPLICATION NUMBER: US/09/755.665
; CURRENT FILING DATE: 2001-08-14
; PRIOR APPLICATION NUMBER: U.S.S.N. 60/174,724
; PRIOR FILING DATE: 2000-01-06
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-755-665-32

Alignment Scores:
Pred. No.: 0.935 Length: 147
Score: 87.00 Matches: 28
Percent Similarity: 43.62% Conservative: 13
Best Local Similarity: 29.79% Mismatches: 20
Query Match: 2.27% Indels: 33
DB: 9 Gaps: 5

US-09-931-733-1 (1-2191) x US-09-755-665-32 (1-147)
QY 230 TCATCTATGAAATTTTCATCTTTGGTATGAATATGAGCATTGCTCGAGT----- 283
DB 3 SerSerLeuLysLeu-----TrpGlyLeuLeuAlaLeuLeu 15
QY 284 -----TGCAGCTATAGTGACCGCAGTTTGTGTGTATGCT----- 319
DB 16 CysLeuLeuCysArgLeu-----CysValTyrSerAsnAsnIleTyrTrp 30
QY 320 -----AGTCCCAATTGAGACATTACTGCATATAAAATG 352
DB 31 ArgGluPheIleLysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
QY 353 ATTATAAGTAAAGGACAGAAAATGTTCTGAAGGGTCATAACTTCCAAGGGTCTCTTACC 412
DB 51 AspValLeuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisMetPhe-Ile-- 69
QY 413 CTTTGACCTCTGGATTGGGGTCCAGTGAGTGCCTTAAT 452
DB 70 -TyrSerLeuTrpPheLysIleGlnArgAlaCysIleAsn 82

RESULT 4
US-10-445-641-6
; Sequence 6, Application US/10445641
; Publication No. US20040053297A1
```


Query Match:	2.24%	Indels:	80
DB:	12	Gaps:	9
US-09-931-733-1 (1-2191) x US-10-424-599-267968 (1-357)			
QY	595	AGGAGTAAGTCAAGAAAAGCAAAGAGGATCTGGAATATGCCATAACATCTCTGGGTG	536
		:	
Db	165	ArgSerArgPheAlaIysArgCysGluSerLeuAsnThrGlyValSerPheLeuTyr-	183
		:	
QY	535	TTTCGTGATAACTTTTACTGCGCTTAGGGT-	506
		:	
Db	184	TrpIleValGlyPheTyrTrpValValSerGlyGlyAspIleLeuLeuGlnAspAlaPro	203
		:	
QY	505	ACTCTTTACGTGGATACATTGAGTTGGCTTTAAGGCAGTAGGTGACAGAAGGATTAAAGG	446
		:	
Db	204	ArgLeuTyr-TrpLeuAlaValPheLeu-	213
		:	
QY	445	CACTCACTGGGAACCCCAATCCAGAGGTCAAAAGGTAGAGAACCTTGGAAAGTTATGACC	386
		:	
Db	214	-----Al	214
		:	
QY	385	CTTCAGAACATTTTC-----TGCGTTTTA-----	361
		:	
Db	214	apheAspValPhePheAlaIlePheCysValValLeuAlaCysLeuIleGlyLeAlaLe	234
		:	
QY	360	-----CTTATAACTATTATATGCAGTAATGTCTCAATGGGACT	320
		:	
Db	234	uCysCysLeuProCysIleIleAlaIleLeuTyrAlaValAlaGlyGlnGluGlyAl	254
		:	
QY	319	AGCATACACACAAACTCGGGTCACCTATAGCGGCAACTCAGGACAATG-----	271
		:	
Db	254	asergluAlaaspLeuSerMetLeuProLysTyrLysPheArgIleLeuSerAspValas	274
		:	
QY	270	-----CTCATATAATTCATACCACAA	251
		:	
Db	274	plysProSerGlyGlyAlaGlySerMetValProIleGluThr-SerSerAlaTyrLeuGlu	294
		:	
QY	250	AGATGAA-----ATTTCATAGATGAAGCAAGAATGCTTAAAAATGTCATTTTCCC	197
		:	
Db	294	uaSngluArgThrLeuLeuLeuGluAspAlaGluCysIleCysLeuCysSerTyrGlu	314
		:	
QY	196	CCACTTTGTGGCAGTG---GCCAACAAATTGAATTTGTCTATCATACAGAAAAAGTCATT	140
		:	
Db	314	uaSpGlyAlaGluLeuHisAlaLeuProCysAsnHisIspheHisSer---SerCysil	333
		:	
QY	139	AGTAGAATATACAAAATTTCTCTGATAAGCAAAAATCTGCGAATAT	94
		:	
Db	333	eVallystrolLeuLysMetAsnAlaThrCysProLeuCysGlyLysTyr	348
		:	

```

RESULT 8
US-09-826-508-16
; Sequence 16, Application US/09826508
; Patent No. US2001002509A1
; GENERAL INFORMATION:
; APPLICANT: Nabil Elshourbagy
; APPLICANT: Lisa Vawter
; TITLE OF INVENTION: G Protein-Coupled Receptor Polypeptides
; TITLE OF INVENTION: and Polynucleotides
; FILE REFERENCE: GP-70744USB
; CURRENT APPLICATION NUMBER: US/09/826.508
; CURRENT FILING DATE: 2001-04-05
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 331
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-826-508-16

Alignment Scores:
Pred. No.: 3.93
Score: 83.00
Percent Similarity: 39.52%
Length: 331
Matches: 29
Conservative: 20

```


US-09-931-7333-1 (1-2191) x US-09-764-860-328 (1-34)

QY 1581 GGGTACAGTAACTATTTTCGGGTGATGTGCACACTGAAATCTCTGACTTCACCAATATG 1522
Db 8 GlyTyrAsnValHisTyrSerGlyAspGlyCysThrGluSerProAspPheThrIleVal 27

QY 1521 GCATATATCCATGTACGAAA 1501
Db 28 GlnTyrIleHisValThrGln 34

RESULT 13

US-10-074-095-328
Sequence 328, Application US/10074095
Publication No. US20030077704A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC008C1
CURRENT APPLICATION NUMBER: US/10/074,095
CURRENT FILING DATE: 2002-02-14
PRIOR APPLICATION NUMBER: 09/764,860
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION NUMBER: 60/179,065
PRIOR FILING DATE: 2000-01-31
PRIOR APPLICATION NUMBER: 60/180,628
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: 60/214,886
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 60/217,487
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11
PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/218,290
PRIOR FILING DATE: 2000-07-14
PRIOR APPLICATION NUMBER: 60/225,757
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/226,868
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/216,647
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: 60/225,267
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/216,880
PRIOR FILING DATE: 2000-07-07
PRIOR APPLICATION NUMBER: 60/225,270
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/251,869
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/235,834
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: 60/234,274
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/234,223
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: 60/228,924
PRIOR FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: 60/224,518
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26
PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20

PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/236,327
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01
PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01
PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08
PRIOR APPLICATION NUMBER: 60/229,509
PRIOR FILING DATE: 2000-09-05
PRIOR APPLICATION NUMBER: 60/236,367
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/237,039
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,038
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/236,370
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: 60/236,802
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,037
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/237,040
PRIOR FILING DATE: 2000-10-02
PRIOR APPLICATION NUMBER: 60/240,960
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/239,935
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/239,937
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: 60/241,787
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/246,474
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/246,532
PRIOR FILING DATE: 2000-11-08
PRIOR APPLICATION NUMBER: 60/249,216
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/249,210
PRIOR FILING DATE: 2000-11-17
PRIOR APPLICATION NUMBER: 60/226,681
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,759
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/225,213
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/227,182
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 60/225,214
PRIOR FILING DATE: 2000-08-14
PRIOR APPLICATION NUMBER: 60/235,836

Mon Mar 29 09:54:58 2004

```

; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08

Query Match: 2.13% Indels: 0
DB: Gaps: 0

US-09-931-733-1 (1-2191) x US-10-074-095-328 (1-34)
QY 1581 GGGTACAGTTAAACATTTTCGGGTGATGTGCACACTGAAATCTCTGACTTCACCAATATG 1522
Db 8 GlyTyrAsnValHisTyrSerGlyAspGlyCysThrGluSerProAspPheThrIleVal 27
QY 1521 GCATATATCCATGTAACGAAA 1501
Db 28 GlnTyrIleHisValThrGln 34

RESULT 14
US-10-212-872-328
; Sequence 328, Application US/10212872
; Publication No. US20030215893A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C2
; CURRENT APPLICATION NUMBER: US/10/212,872
; CURRENT FILING DATE: 2002-08-07
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 328
; LENGTH: 34
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (?)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-212-872-328

Alignment Scores:
Pred. No.: 2,7 Length: 34
Score: 81.00 Matches: 16
Percent Similarity: 74.07% Conservative: 4
Best Local Similarity: 59.26% Mismatches: 7
Query Match: 2.13% Indels: 0
DB: Gaps: 0

US-09-931-733-1 (1-2191) x US-10-212-872-328 (1-34)
QY 1581 GGGTACAGTTAAACATTTTCGGGTGATGTGCACACTGAAATCTCTGACTTCACCAATATG 1522
Db 8 GlyTyrAsnValHisTyrSerGlyAspGlyCysThrGluSerProAspPheThrIleVal 27
QY 1521 GCATATATCCATGTAACGAAA 1501
Db 28 GlnTyrIleHisValThrGln 34

RESULT 15
US-09-771-209-77
; Sequence 77, Application US/09771209
; Patent No. US20020064817A1
; GENERAL INFORMATION:
; APPLICANT: Buck, Linda
; TITLE OF INVENTION: ODORANT RECEPTORS AND USES THEREOF
; FILE REFERENCE: 0575/38586-B/JPW/ADM
; CURRENT APPLICATION NUMBER: US/09/771,209
; CURRENT FILING DATE: 2002-01-26
; PRIOR APPLICATION NUMBER: US 08/129,079
; PRIOR FILING DATE: 1993-10-05
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 77
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Rattus sp. 18
```


US-09-771-209-77

Alignment Scores:

Pred. No.:	8.79	Length:	312
Score:	80.00	Matches:	39
Percent Similarity:	34.48%	Conservative:	21
Best Local Similarity:	22.41%	Mismatches:	39
Query Match:	2.09%	Indels:	75
DB:	9	Gaps:	7

US-09-931-733-1 (1-2191) x US-09-771-209-77 (1-312)

QY	533	GAACACCCAGAGATGTTTATGGCCATATTCAGATCCTCTTTGCTTTTCTTGACTTACT	592
Db	20	GluHisGlnGlnLeuPheAlaLeuPheLeuLeuMetTyr-----LeuThr	35
QY	593	CCTTTTTGTATAAGTTATTTCATATGGTTTCAAAATCCAG---AATCAGCTATTCACCTTC	649
Db	36	ThrPheLeuGlyAsnLeuLeuValValLeuValGlnLeuAspSerHisLeuHisThr	55
QY	650	ATGGCCCTAGTTGGATTAAACAAATCAGCTTCTTTTGCTCTGT-----	694
Db	56	ProMetTyrLeuPheLeuSerAsnLeuSerPheSerAspLeuCysPheSerSerValThr	75
QY	694	-----	694
Db	76	MetLeuLysLeuLeuGlnAsnIleGlnSerGlnValProSerIleSerTyrAlaGlyCys	95
QY	695	-----AGATGGCCTTGGACAGTACATATTAAACAGT	724
Db	96	LeuThrGlnIlePhePheLeuLeuPheGlyTyr-LeuGlyAsnPheLeuLeuValAl	115
QY	725	CTATGGG-----CCCTTCCACTCATGACCACA-----	752
Db	115	aMetAlaTyrAspArgTyrValAlaIleCysPheProLeuHisTyrThrAsnIleMetSe	135
QY	753	-----TCCTGCTTGAGATCTGCTGGGTTAAACACCACATCATCTGCTAG	799
Db	135	rHisLysLeuCysThrCysLeuLeuValPheTyrIleMetThrSerSer-----	152
QY	800	TTGATGCCATTATTACATCTCTGGGGTTACAGCAGCTAAAGGGCACCCCA-----	851
Db	153	-----HisAlaMetMetHisThrLeuLeuAl	161
QY	852	-----GTGTCCTTAACAGAGAAATACACTGCTGAGC	884
Db	161	aAlaArgLeuSerPheCysGluAsnAsnValLeuLeuAsn	174

Search completed: March 28, 2004, 08:34:32

Job time : 149.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 07:46:26 ; Search time 52.5 Seconds
(without alignments)

8028.786 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataattgagagatttta.....tgaggactgaacagagaaa 2191

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-Q=/cgn2_1/uspto_spool/US09931733/runat_25032004_155917_9663/app_query.fasta_1.2375
-DB=PIR_78 -QFMT=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM_ext HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733 @CEN_1_1_57@runat_25032004_155917_9663 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEW TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -XGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR_78.*

1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	98	2.6	175	2	AB0185
2	88	2.3	128	2	S42795
3	87	2.3	331	2	G84981
4	86.5	2.3	3343	2	T42207
5	82.5	2.2	310	2	E23701
6	82.5	2.2	339	2	A47611
7	82.5	2.2	594	2	H48613
8	81.5	2.1	442	2	S50332
9	81.5	2.1	522	2	S09996
10	80	2.1	312	2	G23701
11	80	2.1	478	2	F96842
12	80	2.1	740	2	B84741
13	79.5	2.1	258	2	I56220
14	79	2.1	250	2	AD1956

15	78	2.0	252	2	A34877
16	78	2.0	504	2	C90634
17	78	2.0	670	2	T50312
18	77.5	2.0	475	2	C83452
19	77.5	2.0	950	2	T09076
20	77	2.0	139	2	S24072
21	77	2.0	399	2	S29480
22	77	2.0	420	2	B97304
23	77	2.0	598	2	S55014
24	77	2.0	1226	2	T45788
25	77	2.0	1407	1	BWB1M1
26	76.5	2.0	227	2	F85433
27	76.5	2.0	276	2	T49123
28	76.5	2.0	335	2	H95152
29	76.5	2.0	928	2	F71541
30	76	2.0	203	2	C64019
31	76	2.0	416	2	S68822
32	76	2.0	504	2	C85485
33	76	2.0	510	2	T11943
34	76	2.0	606	2	S35427
35	75.5	2.0	473	2	A42391
36	75.5	2.0	617	2	T23082
37	75.5	2.0	768	2	T18461
38	75.5	2.0	1018	2	JC4211
39	75	2.0	501	2	T11587
40	75	2.0	604	2	AF0187
41	75	2.0	1336	2	T39978
42	75	2.0	2201	2	AH0095
43	74.5	1.9	202	2	T24134
44	74.5	1.9	242	2	C25192
45	74.5	1.9	242	2	T18490

ALIGNMENTS

RESULT 1

AB0185

probable membrane protein YPO1518 [imported] - Yersinia pestis (strain CO92)

C;Species: Yersinia pestis

C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C;Accession: AB0185

R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B. deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001

A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A;Reference number: AB0001; MUID:21470413; PMID:11586360

A;Accession: AB0185

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-175 <XUR>

A;Cross-references: GB:AL590842; PIDN:CA90341.1; PID:gl5979560; GSPDB:GN00175

C;Genetics:

A;Gene: YPO1518

Alignment Scores:

Pred. No.: 0.0745 Length: 175

Score: 98.00 Matches: 32

Percent Similarity: 43.86% Conservative: 18

Best Local Similarity: 28.07% Mismatches: 32

Query Match: 2.56% Indels: 32

DB: 2 Gaps: 8

US-09-931-733-1 (1-2191) x AB0185 (1-175)

QY 479 CCAACTCAATGATTCACCGTAAG-----AGTACCCATAAGCCAGTAAAGATTATCAC 532

Db 39 ProProSerCysGluArgLysGlnGlnSerAsnIleLeuPro-----His 54

QY 533 GAACACCCAGAGATGTTTATGCGCATATTCAGATCTCTTTGCTTTTCTTGACTTACT 592

Db 55 GluTyPro-----ValGlyLeuThrLeuLeuCysPheSerGlyLeI--- 68

C4b-binding protei
probable carniti
hypothetical prote
cytochrome-c oxida
hypothetical prote
envelope protein -
bombesin receptor
probable membrane
NADH2 dehydrogenas
probable disease r
TATA box-binding p
hypothetical prote
hypothetical prote
v-type sodium ATP
hypothetical prote
ycdy protein homol
neurotensin recept
probable carniti
NADH2 dehydrogenas
env polypeptid -
Ca2+-transporting
hypothetical prote
hypothetical prote
neural adhesion pr
glucosyltransferas
probable sideropho
M-factor ABC-type
probable sideropho
hypothetical prote
C-reactive protein
hypothetical prote

	Qy	593	CCTTTTTCGATAAAGTTATTCAATGGTTCCAAATCCAGAATTCACGTATTCACTTCATCG	652
	Db	69	-----TyLThrPhePheValTrpPheAspLeuAspAlaIleIlePheLeuMet	86
	Qy	653	GCCCTAGTTTGGAATAAACAAATCAGCTTCTTTTTGTCTCTAGATGCCCTCG-----ACA	709
	Db	87	SerLeuSerTrpLeu-----PhePheLeuLeuSerLysTrpArgTrpGlnVal	102
	Qy	710	GTACATATTAAACAGTGCTA-----TGGGCCTTTTCCACT-----	742
	Db	103	ValAlaLeuAsnAlaLeuGlyClyValGlyLeuPheTrpPhePheSerSerGlyPheVal	122
	Qy	743	-----CATGACCACATCCTGCTTCAGATCGTCG	772
	Db	123	AsnIleValTrpHisValAsnIleIleValArgAlaLeuIle	136

RESULT 2
S42795
epididymis-specific protein HE3 alpha - human
C:Species: Homo sapiens (man)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 21-Jul-2000
C:Accession: I37455: S42795
R:Kirchhoff, C.; Pera, I.; Rust, W.; Ivell, R.
Mol. Reprod. Dev. 37, 130-137, 1994
A:Title: Major human epididymis-specific gene product, HE3, is the first representative
A:Reference number: I37455; MUID:94235297; PMID:7514008

Alignment Scores:		
Pred. No.:	0.918	Length:
Score:	88.00	Matches:
Percent Similarity:	43.62%	Conservative:
Best Local Similarity:	29.79%	Mismatches:
Query Match:	2.30%	Indels:
DR:	2	Gaps:
		5

```

US-09-931-733-1 (1-2191) x S42795 (1-128)
Qy      230 TCATCTATGAAATTTCACTTTGTGTTAATAATTATGAGCATTTGCTCGAGT-----283
      |||||:::|||||
Db      3 SerSerLeuysile-----TrpGlyIleLeuLeuAlaLeu 15
Qy      284 -----TGCGCCCTATAGTGACCGCAGTTCTGTGTATGCT-----319
      |||||:::|||||
Db      16 CysIleLeuCysargueu-----CysValTyrSerAsnIleTyrTrp 30
Qy      320 -----AGTCCCAATTGAGACATTACTTGCATATAAAATG 352
      |||||:::|||||
Db      31 ArgGluPheIleuysLeuHisTyrLeuSerProSerArgGluPheLysGluTyrLysCys 50
Qy      353 ATTATAGTAAAGGCGAGAAAATGTTCTGAAGGTCATAACTTCCAAGGGTCTCTTACC 412
      ::::|:::|||||
Db      51 AspValIleuMetArgGluLysGluAlaLeuLysGlyLysSerPheHisThrPhe-Ile-- 69
Qy      413 CTTTGCCTCTGAGTTGGGGTTCCCGAGTCAGTCGCCTTAAT 452
      ::::|:::|||||
Db      70 -TyrSerLeuTyrPheLysIleGlnArgAlaCysIleAsn 82

```

```

RESULT 3
G84981
DNA-directed DNA polymerase (EC 2.7.7.7) [imported] - Buchnera sp. (strain APS)
N:Alternate names: DNA polymerase III delta subunit
C:Species: Buchnera sp.
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: G84981
R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.

```

Nature 407, 81-86, 2000
A;Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp.
A;Reference number: A84930; PMID:1093077
A;Accession: G84981
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-331 <STO>
A;Cross-references: GB:AP000398; GSPDB:GN00144
A;Experimental source: strain APS
C;Genetics:
A;Gene: hola; BU445
C;Keywords: nucleotidyltransferase

Alignment Scores:	
Pred. No.:	1.32
Score:	87.00
Percent Similarity:	40.18%
Best Local Similarity:	21.92%
Query Match:	2.28%
DB:	2
Gaps:	12
Indels:	54
Mismatches:	77
Conservative:	40
Matches:	48
Length:	331

US-09-931-733-1 (1-2191) x G84981 (1-331)

Qy 621 AACCATATGAATACTTTTATCAAAAAAGGAGTAAGTCACAGAAAAGCAAGAGGAACTCGG 562
||||| : : : : : ||||| : : : : :
Db 115 AsnHisLeuSerArgPheIleGlnLysAsnLysSerLeuLysGluPheLysAsn----- 132

Qy 561 AATATGGCCATAAACAATCTCTGGGTGTCGTGATAAATCTTTTACTTGCTGGTCTTAGGGTACTC 502
||||| : : : : : ||||| : : : : :
Db 133 -----TyrAsnIleValSerCys-----PheThrProTyrAsnLeuAsn 145

Qy 501 TTT---ACGTGGATACATTGAGTTGGCCTTAAGGCAGTAGGTGAAGAAGGATATTAAAGC 445
||||| : : : : : ||||| : : : : :
Db 146 PheIleAsnTrpIle----- 150

Qy 444 ACTCACTGGGAACCCCATCCAGAGGTCAAGGGTAGAGAACCCCTGGGAAGTTATNAGCC 385
: : : : : : : : : : : : : : : :
Db 151 ---LysTyrGluIleGlnGluLysIleAsnIleGluGlu----- 163

Qy 384 TTCAGAACATTTTCTGCCTTTTACTT-----ATATCATTTTATATGCA 340
: : : : : ||||| : : : : :
Db 164 ---LysAlaPhePheLeuLeuCysLysTyrTyrGluGlyAsnThrLeuPheIleTyrLys 182

Qy 339 GTAATGTCTCAATTTGGGATAGCATACACAAACTCGGGTCACCTATAGGCGGCACTC 280
: : : : : : : : : : : : : : : :
Db 183 IleLeuAspMetLeuPheIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 202

Qy 279 AGGACAATGCTCCAT-----AATTCATACCACAAGATGAATTTTCATA 235
: : : : : : : : : : : : : : : :
Db 203 LysIleIleIleGluPheAspValserProSerTyrTrpIleAsnSerIlePheGln 222

Qy 234 GATGAAGCAAAGAAATGCTTAAAAATGTGC---ATTTTCCCCCACTTTGTGGCACTGGCC 178
: : : : : : : : : : : : : : : :
Db 223 GlyLysThrGluLysSerPheTyrIleLeuAsnIlePhe-----PheLysLysLysTyr 240

Qy 177 AACAAATGTAATTTGTGCTATCATACAAAAGTCATTAGTAGATATATACAAAATTCCT 118
||||| : : : : : ||||| : : : : :
Db 241 AsnProIleIleLeuValArgSerLeuGlnLysAspLeuLeuGluLeuIleHisMetLys 260

Qy 117 GATAAAGCAAAATGTGGCAATATATATACATATATATAAATATGTAGTCATCATATTAC 58
||||| : : : : : ||||| : : : : :
Db 261 ArgGluLysLys-----IleSerIleTyrValMet-----LeuGluLysTyr 274

Qy 57 AGTGACTAAATTTATTAATCAGGCATTATTCTGAATAAAATCTCTCAATATATGTA 1
||||| : : : : : : : : : : : : : : : :
Db 275 Asn-----IlePheValThrArgArgLysPhePheIleLysAlaPheAsnLysIle 291

```

RESULT 4
T42207
breast cancer susceptibility protein BRCA2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 24-Nov-2003
C:Accession: T42207

```

R;McAllister, K.A.; Haugen-Strano, A.; Hagevik, S.; Collins, N.K.; Brownlee, H.; Futreal
submitted to the EMBL Data Library, February 1997
A;Description: Characterization of the mouse and rat homologs of the BRCA2 breast cancer
A;Reference number: Z22073
A;Accession: T42207

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-3343 <MCA>

A;Cross-references: EMBL:U98653; NID:g2443440; PID:g2443441; PIDN:AAB71378.1

A;Experimental source: strain Sprague-Dawley; testes

C;Genetics:

A;Gene: BRCA2

C;Superfamily: breast/ovarian cancer tumor suppressor protein, BRCA2 type

Alignment Scores:
Pred. No.: 1.98 Length: 3343
Score: 86.50 Matches: 49
Percent Similarity: 38.34% Conservative: 25
Best Local Similarity: 25.39% Mismatches: 57
Query Match: 2.27% Indels: 62
DB: 2 Gaps: 9

US-09-931-733-1 (1-2191) x T42207 (1-3343)

```
QY 1482 AATGATATTAACAAATAAGTTTCTTACAAATTCGATAGCTGAAGAGGATCC 1423
Db 1315 AsnAlaTyrLys -----LeuGluAsnSerAspValSerLysSerSer 1328
QY 1422 AACATTAGAATCTTCTTACCAGCAGACTTACATGTGAGTCATTAGAGACTTATATCCC 1363
Db 1329 ThrSerGlyThr -----ValTyrIleAsn 1336
QY 1362 TTAATGATGCTCAATA -----TAAGAGAAGCAATGGCTTCTTCCCTCCCT 1312
Db 1337 LysGlyAspSerAspLeuProPheAlaGluLysGlyAsnIleTyr -----Pro 1353
QY 1311 TTCCTGTGCAGAGTGGCAGTGCAGGAGGAGATCTTGTGAGGAGCATCTATTGCTACC 1252
Db 1354 GluSerCysThrGlnTyrValArgGluGluAsnAlaGlnIleLysGluSerValSerAsp 1373
QY 1251 AATACTTTACTGCATCTCCTCGTTGGGGAATGTCT---TCTATGAATCGCTATTGTT 1195
Db 1374 LeuThrCysLeuGluValMetLysAlaGluGluThrCysHisMetLysSer ----- 1390
QY 1194 TAAATACTACCTGAAGTATTCCATTAATTAATTAATTAATTAATTAATTAATTAATTA 1135
Db 1391 -----SerAspLysGluGlnLeu-Prose 1398
QY 1134 TCAGAAAGTCATTATGAATATGTAATCTTCTTAAATTTATATGAAGGATTTAACAC 1075
Db 1398 rAspLysMetGluGlnAsn -----MetLysGluPheAsnIle 1410
QY 1074 TAGCTGTCTCTTGACCATGACTAGGTACTTTACTTACCTCAATGATTAGATTACCAAGTCT 1015
Db 1410 eSerPhe-----GlnThrAlaSerGlyLysAsnIle 1420
QY 1014 CAGATTTTCTGCATTATCTCTGGATTAACAACTGGATAGTACTGACTCTCAATGACAG 955
Db 1420 eArgValSerLysGluSerLeuAsnLysSerValAsnIleLeuAspGlnGlu ----- 1437
QY 954 CATCGAAGACTTGAATCTCTGGAAATGCATCGAAC 918
Db 1438 -ThrGluAspLeuThrValThrSerAspSerLeuAsn 1449
```

RESULT 5

E23701

olfactory receptor I3 - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 26-Aug-1999

C;Accession: E23701

R;Buck, L.; Axel, R.

Cell 65, 175-187, 1991

A;Title: A novel multigene family may encode odorant receptors: a molecular basis for od

A;Reference number: A23701; MUID:91191556; PMID:1840504

A;Accession: E23701

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-310 <BUC>

A;Cross-references: GB:M64385; NID:g205831; PIDN:AAA41748.1; PID:g205832

C;Superfamily: Olfactory receptor OR14

C;Keywords: G protein-coupled receptor; transmembrane protein

Alignment Scores:
Pred. No.: 4.14 Length: 310
Score: 82.50 Matches: 56
Percent Similarity: 36.82% Conservative: 39
Best Local Similarity: 21.71% Mismatches: 62
Query Match: 2.16% Indels: 101
DB: 2 Gaps: 12

US-09-931-733-1 (1-2191) x E23701 (1-310)

```
QY 515 CCAGTAAAGATTATCAGGAACCCAGAGATGTTTATGGCATATTCAGATCCCTCTTT 574
Db 16 ProIleProGluGluHisGlnHis-----LeuPheTyrAlaLeuPheLeuValMetTyr 33
QY 575 GCITTTTCTTGACTACTCTCTTTTGTATAAAGTTATTTCATATGTTTCAAATCCG-- 631
Db 34 -----LeuThrThrIleLeuGlyAsnLeuIleLeuValValGlnLeu 49
QY 632 AATTACGACTATTCATCTTCATGGCCCTAGTTTGGATTAAACAATCAGCTTCTTTTCTC 691
Db 50 AspSerGlnLeuHisThrProMetTyrLeuPheLeuSerAsnLeuSerPheSerAspLeu 69
QY 692 TGT-----AGA 697
Db 70 CysPheSerSerValThrMetProLysLeuGlnAsnMetArgSerGlnAspThrSer 89
QY 698 TGGCCTTGG-----ACAGTACATATTAACAGTCTATCGGCC----- 734
Db 90 IleProTyrGlyGlyCysLeuAlaGlnThr-TyrPhePheMetValPheGlyAspMetG 109
QY 735 -----TTTCCACTCAT 745
Db 109 uSerPheLeuLeuValAlaMetAlaTyrAspArgTyrValAlaIleCysPheProLeuH 129
QY 746 GACACATCC-----TGCTTCAGATCTGCTGCTGGGTAA 781
Db 129 sTyrThrSerIleMetSerProLysLeuCysThrCysLeuValLeuLeuLeuTrpMetLe 149
QY 782 CACCACATCATCTGCTAGT----- 800
Db 149 uThrThrSerHisAlaMetMetHisThrLeuLeuAlaAlaArgLeuSerPheCysGluAs 169
QY 801 -----TGATGCCATTTTATTTGACATCTCGGGTTACAGCAGCTAA 841
Db 169 nAsnValValLeuAsnPhePheCysAspLeuPheValLeuLeuLysLeuAlaCysSerAs 189
QY 842 AGGCAC-----CCAGTGTCTTAAACAGAGAAAATACACTGCTG-- 881
Db 189 pThrTyrIleAsnGluLeuMetIlePheIleMetSerThrLeuLeuIleIleIleIlePro 209
QY 882 -----AGCAGAGATTAAACAGACTCTCTCACACTACCGAAGT 919
Db 209 ePheLeuIleValMetSerTyrAlaArgIleIleSerSerIleLeuLysValProSerTh 229
QY 920 TCAGTGCAATTTCCAGGATTTTCAAGTCT-----TCGATGCTGCTCATGTT 964
Db 229 rGlnGlyIle-CysLysValPheSerThrCysGlySerHisLeuSerValValSerLeuP 249
QY 965 GAGAGTCCACTATC-----CAGTTTGTATTATCCAGGAATAATGCA 1005
Db 249 heTyrGlyThrIleIleGlyLeuTyrLeuCysProAlaGlyAsnAsnSer 265
```

RESULT 6

A47611

env polyprotein - Rous sarcoma virus (strain SR-A)

C:Species: Rous sarcoma virus
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 07-Feb-1997
C:Accession: A47611
R:Bova, C.A.; Manfredi, J.P.; Swannstrom, R.
Virolgy 152, 343-354, 1986

A:Title: env genes of avian retroviruses: nucleotide sequence and molecular recombinants
A:Reference number: A47611; MUID:86263387; PMID:3014723

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-339 <BOV>

A:Cross-references: GB:M14901

C:Superfamily: type C retrovirus env polyprotein

C:Keywords: polyprotein

Alignment Scores:
Pred. No.: 4.18 Length: 339
Score: 82.50 Matches: 58
Percent Similarity: 36.15% Conservative: 36
Best Local Similarity: 22.31% Mismatches: 86
Query Match: 2.16% Indels: 80
DB: 2 Gaps: 14

US-09-931-733-1 (1-2191) x A47611 (1-339)

```

QY 1136 CTGGAAGCCAGGCTTGTCTAGTATTATTAATGATGAATACATCTCAGTAGTATTAA 1195
Db 5 LeuGluGlnProGly---AsnLeuTrpIleThrTrpAlaAsnArgThrGly----- 20
QY 1196 ACAATAGCGATTTCATA-----GAAGACATTTCCCAACAGGAGATGCAGT 1243
Db 21 ---GlnThrAspPheCysLeuSerThrGlnSerAlaThrSerProPheGlnThrCys--- 38
QY 1244 AAGTATTGGTACCAATAGATGCTCTACCAAGATCTCCTCGCCATGCCACTCTGTG 1303
Db 39 -----LeuIleGlyLeuProSerProIleSerGlu----- 48
QY 1304 CACAGAAAGGAGGGAGAGAACCATGTTGCTCTCTATATTGACATCATTTAAAG 1363
Db 49 -----GlyAspPheLys 52
QY 1364 GGATATATAAGTCTCTAATGACTCACATGTAAGTCTGCTGGTGAAGAAAGTTCTAATGTTG 1423
Db 53 Gly-----TyrValSerAspThrAsnCysThrThrLeuGlyThrAsp----- 66
QY 1424 GAATCTCTTACGATTCGAATTTTGTGAAGAAAAAACTATTGTGTTTATATACATTT 1483
Db 67 ArgLeuValSerSerAlaAspPheThrGlyGlyProAspAsnSerThrThrLeuThrTyr 86
QY 1484 ATGAGATACAAGTGAATTTCTGTATCATGATATATGATATGATGCTGAGTCAGAGATT 1543
Db 87 ArgLysValSerCys-----LeuLeuLysLeuAsnVal 98
QY 1544 TCAGTGTGCATACCCGAAAAATTTTAACATGTATACCATTAAGTAATTTCTCATCCCC 1603
Db 99 SerMetTrpAspGluProGluLeu-----GlnLeuLeuGly 111
QY 1604 CATTTCCCTCACCACATGCTCCAGCCCTTGCCAGAAACCTGTTCTCTCTGACTAA 1663
Db 112 SerGlnSerLeuProAsnIleThrAsnIleAla----- 122
QY 1664 CAGAAACGAAGCTAAAAACACTGTGGGAGGAGTCTCCACATTTCTTCTCT-----ACT 1717
Db 123 -----GlnIleSerGlyLeuThrGlyCysValGlyPheArgProGlnGlyVal 139
QY 1718 CAATTTCTCTGGGAATAGCAATAGAGCAAGC-----CAGCACTAGTCA 1765
Db 140 ProTrpTyrLeuGlyTrpSerArgGlnGluAlaThrArgPheLeuLeuArgHisProSer 159
QY 1766 GCTAACTAAGTCACTCAACCAAGGCTTTTCTCTGTATCTTTGTTGAGACTTCATTT 1825
Db 160 PheSerLysSerThrGluPro-----PheThrValValThrAlaAspArg-HisAs 176

```

```

QY 1826 TCTTAGCGTTTCTGGAGATTAC---AACATCCTCGGTTCCGTTCTGTTGGAACTTATAC 1880
Db 176 nLeupPheMetGlySerGluTyr-CysGlyAlaTyrGlyTyrArgPheTrpAsnMetTyr 195

```

RESULT 7

H48613

env polyprotein precursor - myeloblastosis-associated virus (strain MAV-2(O)p9) (fragme
C:Species: myeloblastosis-associated virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Feb-1998

C:Accession: H48613

R:Joliot, V.; Boroughs, K.; Lasserre, F.; Crochet, J.; Dambrine, G.; Smith, R.E.; Perba
Virolgy 195, 812-819, 1993

A:Title: Pathogenic potential of myeloblastosis-associated virus: implication of env pr
A:Reference number: A48613; MUID:93331743; PMID:8393249

A:Accession: H48613

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-594 <JOL>

C:Superfamily: type C retrovirus env polyprotein

C:Keywords: polyprotein

Alignment Scores:
Pred. No.: 4.47 Length: 594
Score: 82.50 Matches: 58
Percent Similarity: 36.15% Conservative: 36
Best Local Similarity: 22.31% Mismatches: 86
Query Match: 2.16% Indels: 80
DB: 2 Gaps: 14

US-09-931-733-1 (1-2191) x H48613 (1-594)

```

QY 1136 CTGGAAGCCAGGCTTGTCTAGTATTATTAATGATGAATACATCTCAGTAGTATTAA 1195
Db 61 LeuGluGlnProGly---AsnLeuTrpIleThrTrpAlaAsnArgThrGly----- 76
QY 1196 ACAATAGCGATTTCATA-----GAAGACATTTCCCAACAGGAGATGCAGT 1243
Db 77 ---GlnThrAspPheCysLeuSerThrGlnSerAlaThrSerProPheGlnThrCys--- 94
QY 1244 AAGTATTGGTACCAATAGATGCTCTACCAAGATCTCCTCGCCATGCCACTCTGTG 1303
Db 95 -----LeuIleGlyLeuProSerProIleSerGlu----- 104
QY 1304 CACAGAAAGGAGGGAGAGAACCATGTTGCTCTCTATATTGAGCATCATTTAAAG 1363
Db 105 -----GlyAspPheLys 108
QY 1364 GGAATATAAGTCTCTAATGACTCACATGTAAGTCTGCTGGTGAAGAAAGTTCTAATGTTG 1423
Db 109 Gly-----TyrValSerAspThrAsnCysThrThrLeuGlyThrAsp----- 122
QY 1424 GAATCTCTTACGATTCGAATTTTGTGAAGAAAAAACTATTGTGTTTATATACATTT 1483
Db 123 ArgLeuValSerSerAlaAspPheThrGlyGlyProAspAsnSerThrThrLeuThrTyr 142
QY 1484 ATGAGATACAAGTGAATTTCTGTATCATGATATATGATGCTGAGTCAGAGATT 1543
Db 143 ArgLysValSerCys-----LeuLeuLysLeuAsnVal 154
QY 1544 TCAGTGTGCATACCCGAAAAATTTTAACATGTATACCATTAAGTAATTTCTCATCCCC 1603
Db 155 SerMetTrpAspGluProGluLeu-----GlnLeuLeuGly 167
QY 1604 CATTTCCCTCACCACATGCTCCAGCCCTTGCCAGAAACCTGTTCTCTCTGACTAA 1663
Db 168 SerGlnSerLeuProAsnIleThrAsnIleAla----- 178
QY 1664 CAGAAACGAAGCTAAAAACACTGTGGGAGGAGTCTCCACATTTGTTCTCT-----ACT 1717
Db 179 -----GlnIleSerGlyIleThrGlyCysValGlyPheArgProGlnGlyVal 195
QY 1718 CAATTTCTCTGGGAATAGCAATAGAGCAAGC-----CAGCACTAGTCA 1765

```

Db 196 ProTyrLeuGlyTyrSerArgGlnGluAlaThrArgPheLeuLeuArgHisProSer 215
QY 1766 GCTAACTAGTACTCAACCAAGCCCTTTTCTTCTTATCTTTGCAGATCTTCATTT 1825
Db 216 PheSerTyrSerThrGluPro-----PheThrValValThrAlaAspArg-HisAs 232
QY 1826 TCTAGCGTTCTCGGAGATTAC---AACATCCTCGGGTTCCTGCGAACTTTAC 1880
Db 232 nLeuPheMetGlySerGluTyrCysGlyAlaTyrGlyTyrArgPheTrpAsnMetTyr 251
RESULT 8
S50332
NADH2 dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Katharina tunicata mitochondrion
C:Species: mitochondrion Katharina tunicata
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 03-Jun-2002
C:Accession: S50332
R:Boore, J.L.; Brown, W.M.
Genetics 138, 423-443, 1994
A:Title: Complete DNA sequence of the mitochondrial genome of the black chiton, Katharina
A:Reference number: S50327; MUID:95129806; PMID:7828825
A:Accession: S50332
A:Molecule type: DNA
A:Residues: 1-442 <BOO>
A:Cross-references: EMBL:U09810
C:Genetics:
A:Gene: ND4
A:Genome: mitochondrion
A:Genetic code: SGC4
A:Start codon: ATA
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 4
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;

Alignment Scores:
Pred. No.: 5.57 Length: 442
Score: 81.50 Matches: 55
Percent Similarity: 32.06% Conservative: 29
Best Local Similarity: 20.99% Mismatches: 57
Query Match: 2.13% Indels: 121
DB: 2 Gaps: 14

US-09-931-733-1 (1-2191) x S50332 (1-442)

QY 194 TGGGGGAAATGCACATTTTAAAGCATTTCTC-----TTTGCT 229
Db 16 TrpTyrLysMetSerTrpIleSerHisPheMetPhePheMetIleThrLeuAsnPhePhe 35
QY 230 TCATCTATGAATTTTCATCTTTGTGTGATGAATATGGAGC----- 271
Db 36 SerProMetAsnIleProLysIleAsnTyrGlu***PheSerAspAsnPheSerValPro 55
QY 272 ---ATTGCTCTGAGTTGCGCCCTATAGTGCACGCGAGTTTGTGTATGCTAGTCCCAAT 328
Db 56 LeuMetMetLeuSerCysTrp-----IleThrGlyLeuMetIleMetSerSer----- 71
QY 329 TGACACATTACTGCATATAAATGATTATAAGTAAAGGAGGAGAAATGTTCTGAAGGGT 388
Db 72 -----TyrLys-----IleLeuLysAsn 77
QY 389 CATAACTTCAAGGGTTC----- 406
Db 78 AsnAsnLeuValSerPhePheLeuLeuAsnValMetIleLeuAsnPheIleIleMet 97
QY 407 -----TCTACCTTTGACCT 421
Db 98 ValPheThrGlnLysSerLeuPheSerLeuTyrIlePhePheGluAlaSerLeuIlePro 117
QY 422 -----CTGGATTGGGGTCCCGAGTGCAGTGCCTTAATATCCTTCTTCACCT 466
Db 118 ThrLeuIleLeuIleLeuMetTrpGlyTyrGln-----Pro 129
QY 467 ACTGCCTTAAGGCCAACATCAATGATATCCACGTAAAGAGTACCCTAAAGCCAGTAAAGAT 526
Db 130 GluArgLeuGln----- 133

QY 527 TATCAGAAACACCCAGAGATGTTTATGGCCATATTCAGATCCTCTTTGCTTTTCTTGA 586
Db 134 -----AlaGlyMetTyrMetMetIleTyrThrIleLeuGlyAlaLeu----- 147
QY 587 CTTACTCTCTTTTGTATAAGTTATTATCATATGTTTCAAAATCCAGAAATTCAGTATTCAC 646
Db 148 -----ProPheLeuIleAsnIlePhePheIleTyrSer---HisAsnAlaHisLeuAsn 164
QY 647 TTCATGGCCCTAGTT-----TG3ATT 667
Db 165 LeuLeuIleMetMetSerLeuProIleMetProTyrGlnAlaMetIleSerPheTrp 184
QY 668 AAAACAATCAGCTTTCTTTCTCTCTAGATGGCCCTTGGACAGTACATATTAACAGTCTA 727
Db 185 LeuPheIleIleLeuValPheLeuValLysLeuPro-----IleTyrSerPheHisLeu 202
QY 728 TGGGCCCTTTCCACTCAT-----GACCACATCTCTGTTGAGATCTG 769
Db 203 TrpLeuProLysAlaHisValGluAlaProValAlaGlySerMetIleLeuAlaLeu 222
QY 770 CTGCTG 775
Db 223 LeuLeu 224
RESULT 9
S09996
nuclear factor I-A1 - chicken
C:Species: Gallus gallus (chicken)
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 08-Dec-2000
C:Accession: S09996
R:Rupp, R.A.W.; Kruse, U.; Muthaup, G.; Goebel, U.; Beyreuther, K.; Sippel, A.E.
Nucleic Acids Res. 18, 2607-2616, 1990
A:Title: Chicken NF1/TGSCA proteins are encoded by at least three independent genes: NF
A:Reference number: S09995; MUID:90251434; PMID:2339052
A:Accession: S09996
A:Molecule type: mRNA
A:Residues: 1-522 <RUP>
A:Cross-references: EMBL:X51486; NID:g63661; PIDN:CAA35853.1; PID:g63662
C:Superfamily: human nuclear factor I
C:Keywords: DNA binding; transcription regulation

Alignment Scores:

Pred. No.: 5.68 Length: 522
Score: 81.50 Matches: 60
Percent Similarity: 35.71% Conservative: 35
Best Local Similarity: 22.56% Mismatches: 92
Query Match: 2.13% Indels: 79
DB: 2 Gaps: 13

US-09-931-733-1 (1-2191) x S09996 (1-522)

QY 412 CTTTCCACCTCTGGATTGGG-----TTCCAGTGCAGTGCCTTAATATCTCTTCTCACC 465
Db 235 ProIleAlaIleGlyThrGlyProAsnPheSerLeuSerAspLeuGluSerSerTyr 254
QY 466 TACTGCTTAAAGCCCAACTCAATGATATCCACGTAAAGAGTACCCTAAAGCAGT----- 519
Db 255 TyrSerMetSerProGlyAlaMet-----ArgArgSerLeuProSerThrSerThr 272
QY 520 -----AAAAGATTATCAAGAACCCAGAGATGTTTATGGCCATATTCAGATCTC 570
Db 273 SerSerThrLysArgIleLysSerValGluAspGluMetAspSerProGlyGluGluPro 292
QY 571 CTTTGGCTTTTCTTGACTTAC-----TCCTTT 597
Db 293 -----PheTyrThrSerGlnGlyArgSerProGlySerGlySerGlnSerSerGlyTrp 310
QY 598 TTTGATAAAGTTATTATCATATGTTTCAAAATCCAGAAATTCAGCTTATTCATGCGCCT 657
Db 311 HisGluValGluProGlyTyrLeuArgAsnProGluHisArgGlyAlaLeuHisGly--- 329
QY 658 AGTTTGGATTAAACAACATCAGCTTTCTTTGCTCTGTAGATGGCCTTGGACAGTACATAT 717

```

Db      330  -----MetProSerProThrAlaLeu 336
QY      718  TAACAGTCTATGGGCCCTTTCCACTCAGACCATCTGCTGCAGATCTGCTGGG 777
Db      337  LyslySerGluYysSerGlyPheSerProSerProSerGlnThrSerSerLeuGly 356
QY      778  -----TTAACACACATCATCTCTAGTTGATCCATTTATTGACATCTCTGGGGTTAC 831
Db      357  ThrAlaPheThrGlnHisArgProVal ----- 366
QY      832  AGCAGCGTAAAGGCGACCCAGTGTCTTAAACAGAGAAAAATACACTGCTGAGCAGAGGAT 891
Db      367  -----lleThrGly-Pro ----- 370
QY      892  TAAACAGACTCTCTCACTACCGGAAGTTCAAGTTCAGTGCATTTCCAGGATTTCAAGTCTTCG 951
Db      371  ArgAlaSerProHisAlaThrProSerThrLeuHisPhePro--ThrSerProIle-- 388
QY      952  ATGCTGTCTATCTGAGAGTCCACTATCCAGTGTGTTTATCCAGGAATA ----- 1000
Db      389  -----lleGlnProGlyProTyroPheSerHisProAlaIleArgTyroHisPr 405
QY      1001  -ATGCAGAAATCTGAGCATTCGTGAATCTATCAATTCAGTAAAGTAACCTAGTCATGG 1059
Db      405  oGlnGluThrLeuLysGluPheValGlnLeuValCys -----ProAspAlaGl 421
QY      1060  TCAAGAGCAGCTAGTGTAACTCTTCATATATAATTAAAGGAAGATTACATATTC 1119
Db      421  yGlnGln-AlaGlyValGlyPheLeuAsnProAsnGlySerSerGlnGlyLysValH 441
QY      1120  ATAATGACTTCTG 1133
Db      441  lAsnProPheLeu 445

```

RESULT 10

```

G23701
A:Species: Rattus norvegicus (Norway rat)
C:Date: 22-Jan-1993 #sequence_revision 22-Jan-1993 #text_change 26-Aug-1999
C:Accession: G23701
R:Buck, L.; Axel, R.
Cell 65, 175-187, 1991
A:Title: A novel multigene family may encode odorant receptors: a molecular basis for od
A:Reference number: A23701; MUID:91191556; PMID:1840504
A:Accession: G23701
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-312 <BUC>
A:Cross-references: GB:M64387; NID:g205835; PIDN:AAA41750.1; PID:g205836
C:Superfamily: olfactory receptor OR14
C:Keywords: G protein-coupled receptor; transmembrane protein

```

```

Alignment Scores:
Pred. No.: 7.83 Length: 312
Score: 80.00 Matches: 39
Percent Similarity: 34.48% Conservative: 21
Best Local Similarity: 22.41% Mismatches: 39
Query Match: 2.09% Indels: 75
DB: 2 Gaps: 7

```

US-09-931-733-1 (1-2191) x G23701 (1-312)

```

QY      533  GAACACCCAGAGATCTTATGTCATATCCAGATCTCTTTGCTTTTCTTGACTTACT 592
Db      20  GluHisGlnGlnPhePheAlaLeuPheLeuIleMetTyro-----LeuThr 35
QY      593  CTTTCTTTTATGATAAAGTTTATTCATATGTTTCCAAATCCAG--AAATCAGCTTACCTTC 649
Db      36  ThrPheLeuGlyAsnLeuLeuValValLeuValGlnLeuAspSerHisLeuHisThr 55
QY      650  ATGGCCCTAGTTTGGATTAAACATACAGCTTCTTTTGTCTCTGT----- 694

```

```

Db      56  ProMetTyroPheLeuSerAsnLeuSerPheSerAspLeuCysPheSerSerValThr 75
QY      694  ----- 694
Db      76  MetLeuLysLeuLeuGlnAsnIleGlnSerGlnValProSerIleSerTyroAlaGlyCys 95
QY      695  -----AGATGGCTTGGACAGTACATATTAACAGT 724
Db      96  LeuThrGlnIlePhePhePheLeuLeuPheGlyTyro-LeuGlyAsnPheLeuLeuValAl 115
QY      725  CTATGGG-----CCCTTTCACCTCATGACCACA----- 752
Db      115  aMetAlaTyroAspArgTyroValAlaIleCysPheProLeuHisTyroThrAsnIleMetSe 135
QY      753  -----TCCTGCTTGCAGATCTGCTCTGGTTAAACACACATCATCTCTGCTAG 799
Db      135  rHisLysLeuCysThrCysLeuLeuLeuValPheThrIleMetThrSerSer----- 152
QY      800  TTGATGCCATTTATTTGACATCTCTGGGTTACAGCAGCTGCTAAAGGGCACCACA----- 851
Db      153  -----HisAlaMetMetHisThrLeuLeuAl 161
QY      852  -----GTGCTCTTAAACAGAGAAAAATACACCTGCTGAGC 884
Db      161  aAlaArgLeuSerPheCysGluAsnAsnValLeuLeuAsn 174

```

RESULT 11

```

F96842
A:hypothetical protein F23A5.32 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F96842
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Kraykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: F96842
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-478 <STO>
A:Cross-references: GB:AF005173; NID:g6503308; PIDN:AAF14684.1; GSPDB:GN00141
C:Genetics:
A:Gene: F23A5.32
A:Map position: 1

```

```

Alignment Scores:
Pred. No.: 8.23 Length: 478
Score: 80.00 Matches: 46
Percent Similarity: 39.77% Conservative: 24
Best Local Similarity: 26.14% Mismatches: 68
Query Match: 2.09% Indels: 38
DB: 2 Gaps: 10

```

US-09-931-733-1 (1-2191) x F96842 (1-478)

```

QY      1094  AAATTAAAGGAAGGATTTACATATTACATTTCTGATCTGGAAGCCAGGCTTG 1153
Db      165  LysPheLysThr---PheAsnPheSer-----ProAspSerLeuSerHisProSerLeu 181
QY      1154  TCAGTTATTAAATGATGATGATCTCAGTGTAGTATTAAACAATAGCATTCATA 1213
Db      182  Met-----SerLeuSerLeuHisSerTyroPheLeu 191
QY      1214  GAACACATTTCCCAACACGAGGATGAGTAAAGTA-----TTGCTAGCATA 1261
Db      192  GluSerSerHisProLeuArgAsnCysSerAsnLeuArgThrLeuLysLeuLeuSerIle 211

```



```
Db 203 erCys-----LysGly-----AspSerG 209
QY 1389 ATGTAAGTCTGCTGGTAAGAAAGTCTTAATGTTGGAACCTCTTTCAGCTAATCAAAATTTT 1448
Db 209 lyGlyProLeuileCysLysGlyValPheHis---AlaLeuValSerGlyGlyTyrLysC 228
QY 1449 GTAGAAAAAACTATTGTTGTTTATATACATATTATCAGATACACAGTGTAATTCGTGA 1508
Db 228 ysGlyIleSerAsnLysProGlyValTyrThrLeuLeuThrLysLys-----TyrGlnT 246
QY 1509 CATGATATATGCATATTTGGTGAAGTCA 1537
Db 246 hrTrpIleLysSerLysLeuAlaProSer 255

RESULT 14
AD1956
phosphoesterase-related protein alr1199 [similarity] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 15-Sep-2003
C:Accession: AD1956
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1956
A:Molecule type: DNA
A:Cross-references: GB:BA000019; PIDN:BA073156.1; PID:gl1130546; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1199
C:Superfamily: phosphoesterase, M0912 type, phosphoesterase core homology
F:3-72/Domain: phosphoesterase core homology <P>C>

Alignment Scores:
Pred. No.: 9.84 Length: 250
Score: 79.00 Matches: 31
Percent Similarity: 38.4% Conservative: 19
Best Local Similarity: 23.8% Mismatches: 40
Query Match: 2.07% Indels: 40
DB: 2 Gaps: 5

US-09-931-733-1 (1-2191) x AD1956 (1-250)
QY 1680 AAACACTGTGGAGAGTCTCCACATGTTTCTTACTCCATTTTCTCTGGGAATAGCA 1739
Db 68 LysGlyTyrTrpTyrGluGlu-----GlnCysPheIleLeuHisGlyLeuGlyProThrGly 85
QY 1740 GAATAGGACAA-----GCCAGCACTAGTCAGCTAAGTAACTGACTC 1781
Db 86 GluAlaAspAspLeuLeuValGlnTyrGlyGluThrValLysLeu----- 101
QY 1782 AACCAAGGCTTTTTCCTTGTTATCTTTCAGATACATCTTCTTCTAGCGTTTCTGGA 1841
Db 102 -----LeuTrpAspSerValSerArgGlnThrValGlnTrpLeu----- 114
QY 1842 GATTACAACTCTCGGTTCCGTTTCTGGGAACCTTACTGATTTATCTCCCTCCCTACA 1901
Db 115 -----ArgThrLeuAspPheGlyPhePheGluLeuAspCysLeuLeuIleHisGlyThr 132
QY 1902 CAATAAAGCATGATTCCTGCAATTCCTGAAGATCTCAAGATCTGGAATCTGTTGAAAAA 1961
Db 133 ThrValSerIleAspGluAlaLeuThrProAspThrSerProIleGlnMetLeuAspArg 152
QY 1962 ATTTCAGTGAGTACTGTTCTCGATTGTTGTTAAATATGATCTTGTCTTCTTCTTGAAGTC 2021
Db 153 LeuThrArg----- 155
QY 2022 CCAGATCAACAAGGGGACATCAGTATTGT 2051
Db 156 -----MetGlnAlaAsnAsnLeuPheCys 163
```

RESULT 15

```
A34877
C4b-binding protein beta chain precursor - human
N:Alternate names: protein S-binding chain, C4b
C:Species: Homo sapiens (man)
C:Date: 20-Jul-1990 #sequence_revision 20-Jul-1990 #text_change 21-Jul-2000
C:Accession: A47107; A34877; A31117
R:Hillarp, A.; Pardo-Manuel, F.; Ruiz, R.R.; Rodriguez de Cordoba, S.; Dahlbaeck, B.
J. Biol. Chem. 268, 15017-15023, 1993
A:Title: The human C4b-binding protein beta-chain gene.
A:Reference number: A47107; MUID:93315479; PMID:8325877
A:Accession: A47107
A:Molecule type: DNA
A:Residues: 1-252 <H12>
A:Cross-references: GB:L11244; NID:gl179682; PID:g387660; GB:L11245; NID:gl179683; PID:gl179679
R:Hillarp, A.; Dahlbaeck, B.
Proc. Natl. Acad. Sci. U.S.A. 87, 1183-1187, 1990
A:Title: Cloning of cDNA coding for the beta chain of human complement component C4b-bi
A:Reference number: A34877; MUID:90138962; PMID:2300577
A:Accession: A34877
A:Molecule type: mRNA
A:Residues: 1-252 <H1>
A:Cross-references: GB:M29964; NID:gl179678; PIDN:AAB59520.1; PID:gl179679
R:Hillarp, A.; Dahlbaeck, B.
J. Biol. Chem. 263, 12759-12764, 1988
A:Title: Novel subunit in C4b-binding protein required for protein S binding.
A:Reference number: A31117; MUID:88315081; PMID:2970465
A:Accession: A31117
A:Molecule type: protein
A:Residues: 18-20, 'PQ', 23, 26-27, 'L', 29, 'P', 31, 'VD', 34-36, 'F' <H13>
C:Genetics:
A:Gene: GDB:C4BPB
A:Cross-references: GDB:125208; OMIM:120831
A:Map position: 1q32-1q32
C:Superfamily: complement factor H repeat homology
C:Keywords: extracellular protein; plasma
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-252/Product: C4b-binding protein beta chain #status experimental <MAT>
F:23-76/Domain: complement factor H repeat homology <FH01>
F:81-134/Domain: complement factor H repeat homology <FH02>
F:139-191/Domain: complement factor H repeat homology <FH03>

Alignment Scores:
Pred. No.: 12.7 Length: 252
Score: 78.00 Matches: 36
Percent Similarity: 33.83% Conservative: 9
Best Local Similarity: 27.07% Mismatches: 28
Query Match: 2.04% Indels: 60
DB: 2 Gaps: 8

US-09-931-733-1 (1-2191) x A34877 (1-252)
QY 268 GAGCATTCCTGAGTCCCGCTTATAGGTGACGCGCATTTGTGTGTATGTTAGTCCAA 327
Db 21 GluHisCysProGluLeuProVal---AspAsnSerIlePheValala----- 36
QY 328 TTGACACATTACTGCATATAAATGATTATAAGTAAAGGCAGAAAATGTTCTGAAGGG 387
Db 37 -----LysGluValGluGly 41
QY 388 TCATAACTTCCAAGGGTCTCTACCCCTTGACCTCTGATTGGGGTTCCTCCAGTGTGCC 447
Db 42 GlnIleLeuGly----- 45
QY 448 TTAATATCTCTTTCACCTAC---TGCCTTAAGGCCAACTCAATGATTCACGTAAGAG 504
Db 46 -----ThrTyrValCysIleLysGlyTyrHisLeuValGlyLysLysThr 60
QY 505 ---TACCTTAAGCCAGTAAGAATATATCAGAACACCCAGAGATGTT---TATG3CCAT 558
Db 61 LeuPheCysAsnAlaSerLysGluTrpAspSerThrThrThrGluCysArgLeuGlyHis 80
```

QY 559 ATTCAGATCCCTTTTCCTTCTTCTGACTTACTCCTTTTTCATTAAGTTATTCATAG 618
 Db |||||
 81 CysProAspPro-----Val 85

QY 619 GTTTCAAATCCAGATTTCAGCTATTTCATTCATGGCCT 657
 Db ::|||
 86 LeuValAsnGlyGluPheSerSer-----GlyPro 96

Search completed: March 28, 2004, 08:13:47
 Job time : 67.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 05:41:11 ; Search time 29 seconds
(without alignments)
7867.980 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataattgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table:

BLOSUM62
Xgapop 10.0, Xgapext 0.5
Xgapop 10.0, Xgapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgm2 1/USPTO.epool/US09931733/runat_25032004 155916 9635/app_query.fasta.1.2375
-DB=SwissProt_42 -QFMT=fastan -SURFIX=rsp -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733 @CGN 1 1 30 @runat 25032004 155916 9635 -NCPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	2.3	147	1 EP3A HUMAN	Q14507 homo sapien
2	87	2.3	331	1 HOL4 BUCAI	P57520 buchnera ap
3	86.5	2.3	3343	1 BRC2 RAT	Q35923 rattus norv
4	82.5	2.2	310	1 OLFO RAT	P23699 rattus norv
5	81.5	2.1	522	1 NFIA CHICK	E17923 gallus gall
6	80	2.1	312	1 OLFO RAT	P23271 rattus norv
7	80	2.1	331	1 GP18 HUMAN	Q14330 homo sapien
8	80	2.1	365	1 GCST CHITE	Q8kbj9 chlorobium
9	79.5	2.1	258	1 GRAB RAT	P49864 rattus norv
10	79	2.1	408	1 TRUD LEPIN	Q8f8n2 leptospira
11	79	2.1	466	1 SHT HELVI	Q25190 heliothis v
12	78.5	2.1	473	1 PRPS SCHPO	Q13615 schizosacch
13	78.5	2.1	653	1 ST12 ARATH	Q9max3 arabidopsis
14	78	2.0	252	1 C4BB HUMAN	P20851 homo sapien
15	78	2.0	409	1 AMB1 BACST	P37113 bacillus st
16	78	2.0	409	1 DXR SYNPK	Q7u8c3 synecococc
17	78	2.0	504	1 CAIT ECOLI	Q8xa30 escherichia
18	77	2.0	399	1 BRS3_CAVPO	P35371 cavia porce

19	77	2.0	598	1 NUSM PETMA	Q35543 petromyzon
20	77	2.0	1407	1 T2D2 YEAST	P32555 saccharomyc
21	76.5	2.0	643	1 ZP64 MOUSE	Q99ke8 mus musculu
22	76	2.0	203	1 NYFI HAEIN	P44102 haenophilus
23	76	2.0	416	1 NTR2 RAT	O63384 rattus norv
24	76	2.0	457	1 NU4M ARBLI	P34941 arbacia lix
25	75.5	2.0	434	1 GSA CERSY	O74038 cenarchaeum
26	75	2.0	501	1 ALG8 SCHPO	Q10479 schizosacch
27	75	2.0	1336	1 NAM1 SCHPO	P78966 schizosacch
28	74.5	1.9	242	1 CRP1 LIMPO	P06205 limulus pol
29	74.5	2.0	456	1 OST4 HUMAN	P39656 homo sapien
30	74.5	1.9	765	1 PHVA ANASP	Q31c02 anabaena sp
31	74.5	2.0	861	1 TOPI BUCAI	P37371 buchnera ap
32	74.5	2.0	1021	1 CONT RAT	Q63198 rattus norv
33	74.5	2.0	2203	1 CCAD RAT	P27732 rattus norv
34	74	1.9	504	1 CAIT ECOLI	P59333 escherichia
35	74	1.9	504	1 CAIT ECOLI	P31553 escherichia
36	73.5	1.9	242	1 CRP3 LIMPO	P06207 limulus pol
37	73.5	1.9	362	1 RDC1 RAT	O89039 rattus norv
38	73.5	1.9	1018	1 CONT HUMAN	Q12860 homo sapien
39	73	1.9	755	1 DNLI YEAST	P04819 saccharomyc
40	73	1.9	776	1 CME3 BACSU	P39695 bacillus su
41	73	1.9	1070	1 YHV4 YEAST	P38850 saccharomyc
42	73	1.9	3565	1 CSM1 HUMAN	Q96p27 homo sapien
43	72.5	1.9	551	1 COX1 SYNPK	Q06473 synecocyst
44	72.5	1.9	1237	1 DP3A CHLMU	Q9PJ77 chlamydia m
45	72.5	1.9	1433	1 VGLM BUNYW	P04505 bunyanwera

ALIGNMENTS

RESULT 1

ID	EP3A HUMAN	STANDARD;	PRT;	147 AA.
AC	Q14507;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Epididymal secretory protein E3 alpha precursor (HE3 alpha).			
GN	FAM12A OR HE3ALPHA.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Epididymis;			
RX	MEDLINE=94235297; PubMed=7514008;			
RA	Kirchhoff C., Pera I., Rust W., Ivell R.;			
RT	"Major human epididymis-specific gene product, HE3, is the first			
RL	representative of a novel gene family.";			
RL	Mol. Reprod. Dev. 37:130-137(1994).			
RN	[2]			
RP	REVISIONS.			
RA	Kirchhoff C.;			
RL	Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.			
CC	-!- FUNCTION: Possible function in sperm maturation.			
CC	-!- SUBCELLULAR LOCATION: Secreted (Potential).			
CC	-!- TISSUE SPECIFICITY: Epididymis, with predominant expression in the			
CC	corpus region. Moderately expressed in the vas deferens; only low			
CC	levels are detectable in the caput and cauda regions.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X76383; CAA53971.2; -;
DR EMBL; X76385; -; NOT ANNOTATED_CDS.
Genew; HGNC:16978; FAM12A.

ID BRCA2_RAT STANDARD; PRT; 3343 AA.
 AC O35923;
 DT 15-MAR-2004 (Rel. 43, Created)
 DT 15-MAR-2004 (Rel. 43, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Breast cancer type 2 susceptibility protein homolog.
 GN BRCA2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Sprague-Dawley; TISSUE=Testis;
 RX MEDLINE=97384941; PubMed=9242436;
 RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
 RA Collins N.K., Futreal P.A., Bennett L.M., Wiseman R.W.;
 RT "Characterization of the rat and mouse homologues of the BRCA2 breast
 RT cancer susceptibility gene.";
 RL Cancer Res. 57:3121-3125(1997).
 CC -!- FUNCTION: May participate in a pathway associated with the
 CC activation of double-strand break repair and/or homologous
 CC recombination (by similarity).
 CC -!- SUBUNIT: Interacts with RAD51 (By similarity).
 CC -!- TISSUE SPECIFICITY: Highest expression in testis. Also expressed
 CC in spleen, skeletal muscle, thymus, mammary gland, heart, ovary,
 CC prostate, liver, lung, kidney and brain.
 CC -!- SIMILARITY: Contains 7 BRCA2 repeats.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U89653; AAB71378.1; --
 DR PIR; T42207; T42207.
 DR InterPro; IPR02093; BRCA2 repeat.
 DR InterPro; IPR008994; Nucleic_acid_OB.
 DR Pfam; PF00634; BRCA2; 8.
 DR PROSITE; PS50138; BRCA2_REPEAT; 6.
 KW Repeat.
 FT REPEAT 984 1018 BRCA2 1.
 FT REPEAT 1197 1231 BRCA2 2.
 FT REPEAT 1405 1439 BRCA2 3.
 FT REPEAT 1503 1537 BRCA2 4.
 FT REPEAT 1645 1669 BRCA2 5.
 FT REPEAT 1828 1845 BRCA2 6.
 FT REPEAT 1939 1973 BRCA2 7.
 FT REPEAT 2019 2053 BRCA2 8.
 SQ SEQUENCE 3343 AA; 372213 MW; 653DB110D2302A8D CRC64;

 Alignment Scores:
 Pred. No.: 1.37 Length: 3343
 Score: 86.50 Matches: 49
 Percent Similarity: 38.34% Conservative: 25
 Best Local Similarity: 25.39% Mismatches: 57
 Query Match: 2.27% Indels: 62
 DB: 1 Gaps: 9

 US-09-931-733-1 (1-2191) x BRCA2_RAT (1-3343)
 QY 1482 AATGATATTAACCAATAAGTTTTTCTTACAAATTCGATGCTGAAGAGTTC 1423
 Db |||||
 1315 AsnAlaTyrLys-----LeuGluAsnSerAspValSerLysSerSer 1328
 |||||
 QY 1422 AACATTAGAACTTCTTACCAGCAGACTTACATGAGTCATTAGAGACTTATATCCC 1363
 |||||
 Db 1329 ThrSerGlyThr-----ValTyrIleAsn 1336
 |||||
 QY 1362 TTAATGATGCTCAATA-----TAAGAGAAGCAACATGGCTTCTTCTCCCTCT 1312

Db 1337 LysGlyAspSerAspLeuProPheAlaAlaGluLysGlyAsnLysTyr-----Pro 1353
 |||||
 QY 1311 TTCCTGTGCACAGAGTGGCAGTGGCAGGAGAGATCTTGTGAGGACATCTATTGCTACC 1252
 |||||
 Db 1354 GluSerCysThrGlnTyrValArgGluGluAsnAlaGlnIleLysGluSerValSerAsp 1373
 |||||
 QY 1251 AATATTCTTACGATCTCTCTGGTGGGAAATGCT--TCTATGAATCGTATTGTT 1195
 |||||
 Db 1374 LeuThrCysLeuGluValMetLysAlaGluGluThrCysHisMetLysSer----- 1390
 |||||
 QY 1194 TAAATACTACTGAAGTATTTCATCAATATTATAACTGACAGCCTGGGCTTTCAGAG 1135
 |||||
 Db 1391 -----SerAspLysGluGlnLeu-ProSe 1398
 |||||
 QY 1134 TCAGAAAGTCATTATGATATGTAATCTCTCTTAAATTTATATGAGAGTTTACAC 1075
 |||||
 Db 1398 rAspLysMetGluGlnAsn-----MetLysGluPheAsnI 1410
 |||||
 QY 1074 TAGCTGTCTCTTGACCATGACTAGTGTACTTACCTCAATGATTAGATTCAAGAAATGCT 1015
 |||||
 Db 1410 eSerPhe-----GlnThrAlaSerGlyLysAsnI 1420
 |||||
 QY 1014 CAGATTTCTGCATTAATCTCGATATAACAAACGTGGATAGTGCATCTCAACATGACAG 955
 |||||
 Db 1420 eArgValSerLysGluSerLeuAsnLysSerValAsnIleLeuAspGlnGlu----- 1437
 |||||
 QY 954 CATCGAGACTTGAATCTCTGGAATGCACTGAAC 918
 |||||
 Db 1438 -ThrGluAspLeuThrValThrSerAspSerLeuAsn 1449
 |||||
 OLFO_RAT STANDARD; PRT; 310 AA.
 ID OLFO_RAT
 AC P23269;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-JUL-1993 (Rel. 26, Last annotation update)
 DE Olfactory receptor-like protein I3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91191556; PubMed=1840504;
 RA Buck L., Axel R.;
 RT "A novel multigene family may encode odorant receptors: a molecular
 RT basis for odor recognition.";
 RL Cell 65:175-187(1991).
 CC -!- FUNCTION: Putative odorant receptor.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- TISSUE SPECIFICITY: Olfactory epithelium.
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; M64385; AAA41748.1; --
 DR PIR; E23701; E23701.
 DR InterPro; IPR000276; GPCR_Rhodopsin.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PS00237; GPCR_RHODOPSIN.
 DR PROSITE; PS00237; G PROTEIN RECEPT FL 1; 1.
 DR PROSITE; PS50262; G-PROTEIN RECEPT FL 2; 1.
 DR G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Olfaction.
 FT DOMAIN 1 23 EXTRACELLULAR (POTENTIAL).

```

FT TRANSMEM 24 48 1 (POTENTIAL).
FT DOMAIN 49 55 2 (POTENTIAL).
FT TRANSMEM 56 77 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 78 98 3 (POTENTIAL).
FT TRANSMEM 99 118 4 (POTENTIAL).
FT DOMAIN 119 137 5 (POTENTIAL).
FT TRANSMEM 138 156 6 (POTENTIAL).
FT DOMAIN 157 194 7 (POTENTIAL).
FT TRANSMEM 195 217 8 (POTENTIAL).
FT DOMAIN 218 234 9 (POTENTIAL).
FT TRANSMEM 235 258 10 (POTENTIAL).
FT DOMAIN 259 270 11 (POTENTIAL).
FT TRANSMEM 271 290 12 (POTENTIAL).
FT DOMAIN 291 310 13 (POTENTIAL).
FT CARBOHYD 3 3 CARBOHYD (GLCNAC. . .) (POTENTIAL).
FT DISULFID 95 187 BY SIMILARITY.
SQ SEQUENCE 310 AA; 35247 MW; 9688233890205B19 CRC64;

Alignment Scores:
Pred. No.: 2.22 Length: 310
Score: 82.50 Matches: 56
Percent Similarity: 36.82% Conservative: 39
Best Local Similarity: 21.71% Mismatches: 62
Query Match: 2.16% Indels: 101
DB: 1 Gaps: 12

US-09-931-733-1 (1-2191) x OLF0_RAT (1-310)
QY 515 CCAGTAAAGATTATACGACACCCAGAGAGTGTATGGCCATATCCAGATCCTCTTT 574
Db 16 ProileProGluHisGlnHis-----LeuphetyrAlaLeuPheLeuValMetTyr 33
QY 575 GCTTTTCTTGACATCTACTCTTTTGTGATAAGTATTCATATGTTTCAATCAG--- 631
Db 34 -----LeuThrThrLeuGlyAsnLeuLeuIleValLeuValGlnLeu 49
QY 632 AATTCAGCTATTCTACTTCATGGCCCTAGTTGGATTAAACAAACAGCTTCTTTTGTCTC 691
Db 50 AspSerGlnLeuHisThrProMetTyrLeuPheLeuSerAsnLeuSerPheSerAspLeu 69
QY 692 TGT-----AGG 697
Db 70 CysPheSerSerValThrMetProLysLeuLeuGlnAsnMetArgSerGlnAspThrSer 89
QY 698 TGGCCTTGG-----ACATGATATATTAACATCTCTATGGCCC----- 734
Db 90 IleProTyrGlyGlyCysLeuAlaGlnThr-TyrPhePheMetValPheGlyAspMetGl 109
QY 735 -----TTTCCACTCAT 745
Db 109 uSerPheLeuLeuValAlaMetAlaTyrAspArgTyrValAlaIleCysPheProLeuHi 129
QY 746 GACACATCC-----TGCTTGAGATCTGCTGCTGGGTAA 781
Db 129 sTyrThrSerIleMetSerProLysLeuCysThrCysLeuValLeuLeuTyrMetLe 149
QY 782 CACCACATCATCTGCTAGT----- 800
Db 149 uThrThrSerHisAlaMetMetHisThrLeuLeuAlaAlaArgLeuSerPheCysGluAs 169
QY 801 -----TGATGCCATTTATTTGACATCTCTGGGGTTACAGCGCTAA 841
Db 169 nAsnValValLeuAsnPhePheCysAspLeuPheValLeuLeuLysLeuAlaCysSerAs 189
QY 842 AGGCAC-----CCAGTGTCTTAAACAGAGAAAATACATGCTG 881
Db 189 pThrTyrIleAsnGluLeuMetIlePheIleMetSerThrLeuLeuIleIleIleProPh 209
QY 882 -----ACGAGAGGATTAACACAGCTCTCTACACTACCGGAAGT 919
Db 209 ePheLeuLeuValMetSerTyrAlaArgIleSerSerIleLeuLysValProSerTh 229
QY 920 TCAGTGCATTTTCCAGGATTTTCAAGTCT-----TCGATGCTGTCATGTT 964

```

```

Db 229 rGlnGlyIle-CysLysValPheSerThrCysGlySerHisLeuSerValValSerLeuP 249
QY 965 GAGAGTCCACTATC-----CAGCTTGTGTTTATCCAGGAATAATGCA 1005
Db 249 hetyrGlyThrIleIleGlyLeuTyrLeuCysProAlaGlyAsnAsnSer 265

RESULT 5
NFIA_CHICK STANDARD; PRT; 522 AA.
AC PI7923;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Nuclear factor 1 A-type (Nuclear factor 1/A) (NFI-A) (NFI-1/A)
DE (CCAAT-box binding transcription factor) (CTF) (TGGCA-binding
DE protein).
DE NFIA OR NFI-A.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RP TISSUE=Embryo;
RX MEDLINE=90251434; PubMed=2339052;
RA Rupp R.A.W., Kruse U., Multhaup G., Goebel U., Beyreuther K.,
RA Sippl A.E.;
RT "Chicken NFI/TGGCA proteins are encoded by at least three independent
RT genes: NFI-A, NFI-B and NFI-C with homologues in mammalian genomes.";
RL Nucleic Acids Res. 18:2607-2616(1990).
CC -/- FUNCTION: Recognizes and binds the palindromic sequence 5'-
CC TGGCNNNNCCAA-3', present in viral and cellular promoters and in
CC the origin of replication of adenovirus type 2. These proteins are
CC individually capable of activating transcription and replication.
CC -/- SUBCELLULAR LOCATION: Nuclear.
CC -/- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=1;
CC Comment=A number of isoforms are produced;
CC Name=1;
CC IsoId=PI7923-1; Sequence=Displayed;
CC -/- SIMILARITY: Belongs to the CTF/NFI family.
CC -/- SIMILARITY: Contains 1 DWA/MH1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X51486; CAA35853.1; -.
CC PIR; S09996; S09996.
CC TRANSFAC; T00544; -.
CC InterPro; IPR000647; CTF_NFI.
CC InterPro; IPR003619; DwaRfin_A.
CC Pfam; PF00859; CTF_NFI; 1.
CC Pfam; PF03165; MH1; 1.
CC SMART; SM00523; DWA; 1.
CC PROSITE; PS00349; CTF_NFI; 1.
CC Transcription regulation; DNA replication; DNA-binding; Activator;
KW Nuclear protein; Multigene family; Alternative splicing.
FT DOMAIN 67 175 DWA.
SQ SEQUENCE 522 AA; 57473 MW; C51F3E5FE6077740 CRC64;

Alignment Scores:
Pred. No.: 3.22 Length: 522
Score: 81.50 Matches: 60
Percent Similarity: 35.71% Conservative: 35
Best Local Similarity: 22.56% Mismatches: 92

```


QY 695 -----AGATGGCCCTGGACAGTACATATTAACAGT 724
 Db 96 LeuThrGlnIlePhePheLeuPheGlyTyr-LeuGlyAanPheLeuLeuValAl 115
 QY 725 CTATGGG-----CCCTTCCACTCATGACCACA----- 752
 Db 115 aMetAlaTyrAspArgTyrValAlaIleCysPheProLeuHisTyrThrAsnIleMetSe 135
 QY 753 -----TCCTGCTTCAGATCGCTGGGTAAACACACATCATCTGCTAG 799
 Db 135 rHisLysLeuCysThrCysLeuLeuValPheTrpIleMetThrSerSer----- 152
 QY 800 TTGATGCCATTATTGATCATCTGGGTACAGCAGCTAAAGGCGCCCA----- 851
 Db 153 -----HisAlaMetMetHisThrLeuLeuAl 161
 QY 852 -----GTGTCCTTAACAGAGAAAATACATGCTGAGC 884
 Db 161 aAlaArgIleSerPheCysGluAanAsnValLeuLeuAan 174
 RESULT 7
 ID GP18 HUMAN STANDARD; PRT; 331 AA.
 AC Q14330; Q96H16;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable G-protein-coupled receptor GPR18.
 GN GPR18 OR GPCR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97349114; PubMed=9205118;
 RA Cantz H., Murakoa A., Yang Y.-K., Samuelson L.C., Zimmerman E.M.,
 RA Cook H., Yamada T.;
 RA "Cloning and chromosomal localization of a gene (GPR18) encoding a
 RT novel seven transmembrane receptor highly expressed in spleen and
 RT testis";
 RL Genomics 42:462-466 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lymph;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalish D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 CC 1- FUNCTION: ORPHAN RECEPTOR. COULD BE A CHEMOKINE RECEPTOR.
 CC 1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC 1- TISSUE SPECIFICITY: MOST ABUNDANT IN TESTIS AND SPLEEN.
 CC 1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; L42324; AAB65819.1; --
 DR EMBL; BC008569; AAB08569.1; --
 DR Genbank; HGNC:4472; GPR18.
 DR MIN; 602042; --
 DR GO; GO:0016021; C: integral to membrane; NAS.
 DR GO; GO:0004930; F: G-protein coupled receptor activity; NAS.
 DR GO; GO:0007186; P: G-protein coupled receptor protein signalin. .; NAS.
 DR InterPro; IPR000276; GPCR_Rhodopsn.
 DR Pfam; PF00001; 7tm_1; 1.
 DR PRINTS; PR00237; GPCRHOPOFSN.
 DR PROSITE; PS00237; G-PROTEIN RECP F1_1; 1.
 DR PROSITE; PS0262; G-PROTEIN RECP F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein.
 FT DOMAIN 1 26
 FT TRANSMEM 27 47
 FT DOMAIN 48 56
 FT TRANSMEM 57 77
 FT DOMAIN 78 95
 FT TRANSMEM 96 116
 FT DOMAIN 117 138
 FT TRANSMEM 139 159
 FT DOMAIN 160 191
 FT TRANSMEM 192 212
 FT DOMAIN 213 232
 FT TRANSMEM 233 253
 FT DOMAIN 254 268
 FT TRANSMEM 269 289
 FT DOMAIN 290 331
 FT CARBOHYD 14 14
 FT CONFLICT 12 12
 FT CONFLICT 86 86
 FT CONFLICT 97 97
 FT CONFLICT 310 310
 FT CONFLICT 331 AA; 38061 MW; 95C9C33BE25D7606 CRC64;
 SQ SEQUENCE
 Alignment Scores:
 Pred. No.: 4.25 Length: 331
 Score: 80.00 Matches: 28
 Percent Similarity: 39.52% Conservative: 21
 Best Local Similarity: 22.58% Mismatches: 47
 Query Match: 2.10% Indels: 28
 DB: 1 Gaps: 5
 US-09-931-733-1 (1-2191) x GP18_HUMAN (1-331)
 QY 603 ATCAAAAAGAGTAAAGTCAAGAAAACAAAGAGGATCTGGAATATGCCATAAACATC 544
 Db 222 LeuLysProLysValGluGluSerLeuArgIleIleThrLeuLeuValGlnVal 241
 QY 543 TCTGGGTGTCGTGATATCTTTTACTGGTACTCTTTTACGTGGATACATGA 484
 Db 242 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 255
 QY 483 GTTGGCCCTTAAGGCGAGTAAAGGATATAAGGCACCTCACTGGGAACCCCAATCC 424
 Db 256 -----MetLeuGlyThrGlyGluAsnSerTyr----- 264
 QY 423 AGAGTCAAGAGGTAGAGAACCTTGGAGTATGACCCCTTCAGAACATTTTCT3CCTT 364
 Db 265 -----AsnProTrpGly-----AlaPheThrThrPheLeuMetAsn 276
 QY 363 TTA-----CTTATAATCATTTTATATGTCAGTAATGTCTCAATTCGGAGTACATAC 313
 Db 277 LeuSerThrCysLeuAspValIleLeuTyrIleValSerLysGlnPheGlnAlaArg 296
 QY 312 ACACAACTGGCGTCCACTATAGGGCGCAACTCAGGACAATGCTCCATAATTCATACAC 253


```

Db      297 ValileSerValMetLeuTyrArgAsnTyrLeuArgSerLeuArgArgLysSerPheArg 316
QY      252 AAAGATGAATT 241
Db      317 SerGlySerLeu 320

RESULT 8
GCST_CHLITE
ID_GCST_CHLITE STANDARD; PRT; 365 AA.
AC Q8XB39;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Aminomethyltransferase (EC 2.1.2.10) (Glycine cleavage system T
DE protein)
DE GCVT OR Ctl788.
GN Chlorobium tepidum.
OS Bacteria; Chlorobi; Chlorobia; Chlorobiales; Chlorobiaceae;
OC Chlorobium.
OX NCBI_TaxID=1097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TL5 / ATCC 49652 / DSM 12025;
RX MEDLINE=22103685; PubMed=1203901;
RA Eisen J.A., Nelson K.E., Paulsen I.T., Heidelberg J.F., Wu M.,
RA Dodson R.J., Deboy R., Gwinn M.L., Nelson W.C., Haft D.H.,
RA Hickey E.K., Peterson J.D., Durkin A.S., Kolonay J.L., Yang F.,
RA Holt I., Umayam L.A., Mason T., Brenner M., Shea T.P., Parksey D.,
RA Nierman W.C., Feldblyum T.V., Hansen C.I., Craven M.B., Radune D.,
RA Venter J.C., Tettelin H., Bryant D.A., Fraser C.M.;
RA "The complete genome sequence of Chlorobium tepidum TLS, a
RT photosynthetic, anaerobic, green-sulfur bacterium.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:9509-9514 (2002).
CC -1- FUNCTION: The glycine cleavage system catalyzes the degradation of
CC glycine (By similarity).
CC -1- CATALYTIC ACTIVITY: (6S)-tetrahydrofolate + S-
CC aminomethyldihydropolypylprotein = (6R)-5,10-
CC methylenetetrahydrofolate + NH(3) + dihydropolypylprotein.
CC -1- SUBUNIT: The glycine cleavage system is composed of four proteins:
CC P, T, L and H (by similarity).
CC -1- SIMILARITY: Belongs to the gcvT family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE012931; AAM73009.1; -.
CC TIGR; Ctl788;
CC HAMAP; MF_00259; -.
CC InterPro; IPR006222; GCV_T.
CC InterPro; IPR006223; GcvT.
CC Pfam; PF01571; GCV_T; 1.
CC TIGRFAMs; TIGR00528; gcvT; 1.
KW Transferase; Aminotransferase; Complete proteome.
SQ SEQUENCE 365 AA; 40139 MW; 5B627C302FF7C86A CRC64;

Alignment Scores:
Pred. No.: 4.34 Length: 365
Score: 80.00 Matches: 27
Percent Similarity: 40.23% Conservative: 8
Best Local Similarity: 31.03% Mismatches: 34
Query Match: 2.10% Indels: 18
DB: 1 Gaps: 3

US-09-931-733-1 (1-2191) x GCST_CHLITE (1-365)
QY 572 AGAGGATCTGGAATATGGCCATAAATCATCTCTGGGTTCGTGATTAATCTTTTACTGGCT 513

```

```

Db      217 LysSerAspGlyIleGlnPro---IleGlyLeuGlyAlaArgAspThrLeuArgLeuGlu 235
QY      512 TTAGGCTACTCTTTTACCTGGATAC-----ATTGAGTTGGCC 477
Db      236 MetGlyTyrSerLeuTyrGlyHisGluIleGluArgAspValAsnProLeuGluAlaArg 255
QY      476 TTAAGGCAGTAGGTGAAGAAGGATATTAAAGGCCTACTCTGGG----- 435
Db      256 LeuLysTrpValLysLeuAsnLysProAsnPheIleGlyLysGlnAlaCysGluGln 275
QY      434 -----AACCCCAATCCAGAGGTCAAAGGGTAGAGAACCCCTTGGAAGCTTATGFCCT 384
Db      276 ValGluIleAsnProArgLysSerValValGlyPheSerLeuGluGlyArgAlaIlePro 295
QY      383 TCAGAACATTTTCTGCTTT 363
Db      296 ArgGlnHisPheLysValTyr 302

RESULT 9
GRK_RAT
ID_GRK_RAT STANDARD; PRT; 258 AA.
AC P49864;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Granzyme K precursor (EC 3.4.21.-) (NK-tryptase-2) (NK-TRYP-2).
GN GZMK.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 26-58.
RC STRAIN=Fischer 344; TISSUE=Lymphocytes;
RX MEDLINE=94179809; PubMed=8133042;
RA Savers T.J., Wiltout T.A., Smyth M.J., Ottaway K.S., Pilaro A.M.,
RA Sowder R., Henderson L.E., Sprenger H., Lloyd A.R.;
RT "Purification and cloning of a novel serine protease, RNK-TRYP-2,
RT from the granules of a rat NK cell leukemia.";
RL J. Immunol. 152:2289-2297 (1994).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic granules.
CC -1- TISSUE SPECIFICITY: Spleen, lungs and liver nonparenchymal cells.
CC -1- SIMILARITY: Belongs to peptidase family S1. Granzyme subfamily.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L19694; AAA42057.1; -.
CC PIR; I56220; I56220.
CC HSSP; P20160; 1A7S.
CC MEROPS; S01.146; -.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001314; Peptidase_S1A.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_SPC; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Zymogen; signal.
FT SIGNAL 1 23 POTENTIAL.
FT PROPEP 24 25 ACTIVATION PEPTIDE.
FT CHAIN 26 258 GRANZYME K.
FT ACT_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 208 208 CHARGE RELAY SYSTEM (BY SIMILARITY).

```


interacts with anticoagulant protein S and with serum amyloid P component. The beta chain binds protein S.

-!- SUBUNIT: disulfide-linked complex of C4BP alpha and beta chains of 3 possible sorts: a 570 kDa complex of 7 alpha chains and 1 beta chain, a 530 kDa homodimer of alpha chains or a 500 kDa complex of 6 alpha chains and 1 beta chain. The central body of the alpha chain homopolymer supports tentacles, each with the binding site for C4b at the end.

-!- SIMILARITY: TO C4BP ALPHA CHAIN AND TO PIG APOLIPOPROTEIN R.

-!- SIMILARITY: Contains 3 Sushi (SCR) domains.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; L11244; AAA35615.1; -
 EMBL; L11245; AAA35616.1; -
 EMBL; M29964; AAB59520.1; -
 PIR; A47107; A34877.
 HSP; P10998; LVVD.
 Genew; HGNC:1328; C4BPB.
 MIM; 120831; -
 GO; GO:0007596; P: blood coagulation; TAS.
 InterPro; IPR000436; Sushi_SCR_CCP.
 Pfam; PF00084; sushi; 3.
 SMART; SM00032; CCP; 3.
 Complement pathway; Plasma; Glycoprotein; Repeat; Sushi; Signal; Polymorphism.

FT SIGNAL 1 17
 FT CHAIN 18 252 C4B-BINDING PROTEIN BETA CHAIN.
 FT DOMAIN 22 77 SUSHI 1.
 FT DOMAIN 80 135 SUSHI 2.
 FT DOMAIN 138 192 SUSHI 3.
 FT DISULFID 23 63 BY SIMILARITY.
 FT DISULFID 49 76 BY SIMILARITY.
 FT DISULFID 81 121 BY SIMILARITY.
 FT DISULFID 107 134 BY SIMILARITY.
 FT DISULFID 139 179 BY SIMILARITY.
 FT DISULFID 165 191 BY SIMILARITY.
 FT DISULFID 202 202 INTERCHAIN (WITH ALPHA CHAIN) (POTENTIAL).
 FT DISULFID 216 216 INTERCHAIN (WITH ALPHA CHAIN) (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 117 117 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 154 154 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARIANT 198 198 P > S (in dbSNP:1803226).
 FT /FTID=VAR_012039.
 SQ SEQUENCE 252 AA; 28357 MW; 0F6CC64067C5E27 CRC64;

Alignment Scores:
 Pred. No.: 6.62 Length: 252
 Score: 78.00 Matches: 36
 Percent Similarity: 33.83% Conservative: 9
 Best Local Similarity: 27.07% Mismatches: 28
 Query Match: 2.04% Indels: 60
 DB: 1 Gaps: 8

US-09-931-733-1 (1-2191) x C4BB_HUMAN (1-252)

QY 268 GAGCATTTCTGAGTTCGCGCTATAGTACCCAGTTTGTGTATGATGTCACCA 327
 |||||
 Db 21 GluHisCysProGluLeuProProVal---AspAsnSerIlePheValAla----- 36
 |||||
 QY 328 TTGAGACATTACTGCATATATAAATGATTATAGTAAAGGCGAGAAATGTTCTGAAGG 387
 |||||
 Db 37 -----LysGluValGluGly 41

QY 388 TCATTAACCTCCAAAGGGTTCTCTACCTTTGACCTCTGAGTTGGGGTTCCCAAGTGAATGCC 447
 |||||
 Db 42 GlnIleLeuGly----- 45
 QY 448 TTAATATCTCTCTTCCACCTAC---TGCTTAAGGCCAACTCAATGCTATATCCACGTAAAGAG 504
 |||||
 Db 46 -----ThrTyValCysIleIysGlyTyHisLeuValGlyIleLysThr 60
 QY 505 ---TACCCTAAAGCCAGTAAAGATTATACGAAACACCCAGAGATGTT---TATGSCCAT 558
 |||||
 Db 61 LeuPheCysAsnAlaSerLysGluTyrAspAsnThrThrThrGluCysArgLeuGlyHis 80
 QY 559 ATTCAGATCTCTCTTCTTTCTTTCTTCTGACTTACTCTCTTTTGTATAAAGTTTATTCATATG 618
 |||||
 Db 81 CysProAspPro-----Val 85
 QY 619 GTTCAATCCAGAAATTCAGCTATTCTTCTCATGCGCCT 657
 |||||
 Db 86 LeuValAsnGlyGluPheSerSerSer-----GlyPro 96

RESULT 15
 AMB1_BACST
 ID AMB1_BACST STANDARD; PRT; 409 AA.
 AC P37113; P94345;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE N-carbamoyl-L-amino acid hydrolase (EC 3.5.1.87) (L-carbamoylase).
 GN AMAB.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
 OX NCBI_TaxID:1422;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIB 8224;
 RX MEDLINE=97176409; PubMed=9023955;
 RA Batisse N., Weigel P., Lecocq M., Sakanyan V.;
 RT "Two amino acid amidohydrolase genes encoding L-stereospecific
 carbamoylase and aminoacylase are organized in a common operon in
 Bacillus stearothermophilus."
 RT Appl. Environ. Microbiol. 63:763-766 (1997).
 RN [2]
 RP SEQUENCE OF 60-409 FROM N.A.
 RC STRAIN=NCIB 8224;
 RX MEDLINE=94113715; PubMed=8285691;
 RA Sakanyan V., Desmarez L., Legrain C., Charlier D.R.M., Mett I.,
 RA Kochikyan A., Savchenko A., Boyen A., Falmagne P., Pirard A.,
 RA Glansdorff N.;
 RT "Gene cloning, sequence analysis, purification, and characterization
 of a thermostable aminoacylase from Bacillus stearothermophilus."
 RT Appl. Environ. Microbiol. 59:3878-3888 (1993).
 RL -!- CATALYTIC ACTIVITY: N-carbamoyl-L-2-amino acid (a 2-ureido
 carboxylate) + H(2)O = L-2-amino acid + NH(3) + CO(2).
 CC -!- COFACTOR: Manganese or nickel or cobalt.
 CC -!- SIMILARITY: Belongs to peptidase family M40.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL; Y08752; CAA69999.1; -
 EMBL; X74289; CAA52341.1; -
 DR MESOPS; M40.UNW; -
 DR InterPro; IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; 1.
 KW Hydrolase.
 FT CONFLICT 60 62
 FT SEQUENCE 409 AA; 44166 MW; 9A91F767C9AEEFBB CRC64;

Alignment Scores:

Pred. No.:	7.38	Length:	409
Score:	78.00	Matches:	36
Percent Similarity:	36.00%	Conservative:	18
Best Local Similarity:	24.00%	Mismatches:	40
Query Match:	2.05%	Indels:	56
DB:	1	Gaps:	6

US-09-931-733-1 (1-2191) x AMB1_BACST (1-409)

Qy	1509	GTACGAAATACACTTCTATCTC-----ATAAATGTATATAAACAATAAGTTT	1456
Db	262	ValGlyGlnLeuHisValTyrProGlyGlyIleAsnValIleProGluArgValGluPhe	281
Qy	1455	TTCCTACAA-----AATTCGATAGCT	1435
Db	282	ValLeuAspLeuArgAspLeuLysAlaGluValArgAspGlnValTrpLysAlaIleAla	301
Qy	1434	GAAGAGGAGTCCAAACATTAGAACTTCTTACCAGCAGACTTACATGTGAGTCATTAGAG	1375
Db	302	ValArgAlaGluThrIleAlaLysGluArgAsnValArgLeuThrGluArgLeuGln	321
Qy	1374	ACTTATATTCCTTAAATGATGCTCAATATAAGAG-----AAG	1336
Db	322	GluMetAlaProValLeuCysSerGluValValLysGlnAlaAlaGluArgAlaCysLys	341
Qy	1335	CAACATGGCTTCTCTCCCTCCCTTC-----	1309
Db	342	GlnLeuGlyTyr-----ProProPheTrpLeuProSerGlyAlaAlaHisAspGlyVal	359
Qy	1308	-----CTGTGC-----	1303
Db	360	GlnLeuAlaProIleCysProIleGlyMetIlePheValArgSerGlnAspGlyValSer	379
Qy	1302	-----ACAGATGGCAGTGGCAGGAGAGATCTTGTAGGAGCATCTATTGCTACC	1252
Db	380	HisSerProAlaGluTrpSerThrLysGluAspCysAlaValGlyAlaGluValLeuTyr	399
Qy	1251	AATACTTTACTGCATCTCCGTTGGGAA	1222
Db	400	HisThrValTrpGlnLeuAlaGlnGlyGlu	409

Search completed: March 28, 2004, 08:06:16

Job time : 44 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: March 28, 2004, 07:42:51 ; Search time 164.5 Seconds

(without alignments)

8404.868 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 3825

Sequence: 1 tacataatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-O=/cpn2.1/USPTO.spool/US09931733/runat_25032004.155917.9648/app_query.fasta.1.2375
-DB=SPREMBL_25 -QFWT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=bl6sum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09931733 @CGN 1.1 187 @runat_25032004.155917.9648 -NCPU=3
-NO_WMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SPREMBL_25.*

1: sp archaea.*
2: sp bacteria.*
3: sp fungi.*
4: sp human.*
5: sp invertebrate.*
6: sp mammal.*
7: sp mhc.*
8: sp organelle.*
9: sp phase.*
10: sp plant.*
11: sp rodent.*
12: sp virus.*
13: sp vertebrate.*
14: sp unclassified.*
15: sp rvirus.*
16: sp bacteriap.*
17: sp archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	2.7	3290	10 Q9SH73	Q9sh73 arabidopsis

2	98	2.6	175	16	Q8ZG02
3	98	2.6	220	16	Q8CKX8
4	94	2.5	3135	5	Q813B5
5	89.5	2.3	906	3	Q870P3
6	86.5	2.3	374	6	Q8HZR7
7	86.5	2.3	3343	11	Q35923
8	85	2.2	409	2	Q8QG55
9	84.5	2.2	230	16	Q8PAM4
10	83.5	2.2	507	5	Q9GTJ5
11	83	2.2	333	4	Q9H2L2
12	83	2.2	852	5	Q81B40
13	83	2.2	1208	16	Q7VRE9
14	82.5	2.2	396	15	Q85483
15	82.5	2.2	446	8	Q8HD85
16	82.5	2.2	600	15	Q83129
17	82	2.2	648	16	Q82222
18	82	2.1	676	3	Q9P7E5
19	82	2.2	837	13	Q7T1T7
20	82	2.2	921	5	Q7YZ56
21	81.5	2.1	479	11	Q7TSF9
22	81.5	2.1	582	5	Q8N515
23	81	2.1	205	16	Q893A7
24	81	2.1	866	11	Q80XL8
25	80.5	2.1	389	16	Q83E71
26	80.5	2.1	443	8	Q34844
27	80	2.1	302	5	Q81471
28	80	2.1	478	10	Q8SAGA
29	80	2.1	740	10	Q49325
30	80	2.1	1419	5	Q813G5
31	79.5	2.1	272	4	Q95992
32	79.5	2.1	330	16	Q8E4H8
33	79.5	2.1	330	16	Q8DYW9
34	79	2.1	250	16	Q8YXL3
35	79	2.1	310	11	Q7TRX5
36	79	2.1	335	3	Q9P6N1
37	78.5	2.1	396	2	Q9EVD2
38	78.5	2.1	2954	12	Q96898
39	78	2.0	251	4	Q9BS25
40	78	2.0	292	16	Q814H9
41	78	2.0	312	11	Q8VGR6
42	78	2.0	314	16	Q889G9
43	78	2.0	333	8	Q9MGE0
44	78	2.0	335	12	Q91HQ2
45	78	2.0	409	16	Q7U8C3

ALIGNMENTS

RESULT 1
Q9SH73
ID Q9SH73 PRELIMINARY; PRT; 3290 AA.
AC Q9SH73
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE F22C12.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Khan S., Brooks S., Buehler E., Chao Q., Dunn P., Kim C.,
RA Walker M., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
RT "Genomic sequence for Arabidopsis thaliana BAC F22C12 from chromosome
I.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007764; AAF24584.1; -
DR GO; GO:0005524; F:ATP binding; IEA.

RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., Mclean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrell B.G.,
 RT "Sequence of Plasmidium falci-parum chromosomes 1, 3-9 and 13";
 RL Nature 419:527-531(2002).
 DR EMBL; AL929355; CAD51718.1; -
 KW Hypothetical protein.
 SQ SEQUENCE 3135 AA; 372458 MW; 40548E3A4123545F CRC64;

Alignment Scores:
 Pred. No.: 0.298 Length: 3135
 Score: 94.00 Matches: 62
 Percent Similarity: 36.08% Conservative: 43
 Best Local Similarity: 21.31% Mismatches: 101
 Query Match: 2.46% Indels: 85
 DB: 5 Gaps: 15

US-09-931-733-1 (1-2191) x Q813B5 (1-3135)

QY 1 TACATAATGAGAGATTATATAGAAATAGCCTGATTAATAATAAATAGTCACTGTA 60
 Db 622 TyrPheValAspLys--TyrSerIleSerCysLeuIleTyrMetAsnIleTyrThrGlu 640
 QY 61 ATAATGATGACTACATAATATATA-----TATGTATAT-----ATA 96
 Db 641 GluSerIleIleLeuLysTyrIleIleThrIleLeuPheTyrIleHisCysAspLeuPhe 660
 QY 97 TTCGCACATTTTTCCTTATCAGAAATTTTGTATATCTACTATGACTTTCTGTATG 156
 Db 661 PheLysIleSerTyrLeuGluAsnMet-----Tyr 671
 QY 157 ATAGCAAAATACAAATGTTGGCCACTGCCACAAAGTGGGGGAAATGACATTTTAA 216
 Db 672 IleSerLysThrAsnIleSerLeuTyrCysTyrLys-----GluSerHisHisIleSer 689
 QY 217 GCATTT-----CTTTGCTTCATCTATGAA 240
 Db 690 LeuTyrAsnTyrIleLysLysIleSerSerIlePheLeuPheCysPheIleTyrSer 709
 QY 241 AATTTTC---ATCTTT---GTGTATGAATATGAGACATTTCTCCTGAGTCCGCGCT 291
 Db 710 GluPheAspIleTyrGlnAsnValIleTyrPheProLysHisLeuHisAsnIle----- 727
 QY 292 ATAGGTGACCGCAGTTTGTGTATGCTAGTCCCAATGACACATTAAGTATATAAAT 351
 Db 728 -----ThrHisIleLeuHisPheCys----- 734
 QY 352 GATTAATAAGTAAGGACGAGAAATGTTCTGAAGGTCATTAATCTCCAGGGTCTCTAC 411
 Db 735 AspGluProTyrIleSerSerSerCysIleGlnSer----- 747
 QY 412 CCTTTGACCTCGAATGGGGTTCACAGTGAGTCCCTTAATATCCTTCTTCCACCTAC--- 468
 Db 748 -----IleTyrPhePheLeuTyrLys 754
 QY 469 -----TGCCCTTAAGGCAACTCAATGATATCCAGTAAGAGTACCCTTAAGCCAGTAAA 522
 Db 755 TyrIleCysTyrHisPhePheSerIlePheIleGlnAsnIleLeuTyrLysIleMetHis 774
 QY 523 AGATATACAGAACACCCAGAGATGTTTATGGCCATATCCAGATCCTCTT----- 573
 Db 775 SerThrSerProIleArgHisGluGluAsnAspHisIlePheTyrProLeuSerValLeu 794
 QY 574 -----TGCTTTTCTTGACTTACTCCTTTTGTATGAAGTTATTCATATGTT 621
 Db 795 LysIleMetLysCysPheAsnTyrTyrTyrAspTyrPheSerAsnIlePheIleArgGlu 814
 QY 622 TCAATCCAGAAATCAGCTATTCATCTCAGTGGCCCTAGTTTGGATTAACCAATCAGCTT 681

Db 814 uAnIleLysMetAsnLeuIleLeuLysGluArgCysLeuIleMetLysAspIleGluAs 834
 QY 682 TCTTTTCTCTCTAGATGGCTTGGACAGTACATATT-----AACAGTCTATGGGCCCT 735
 Db 834 nIleLysIleLeuIleMetProLeuThrTyrAsnIleValCysAsnAsnMetTyrPheNty 854
 QY 736 TTCC-----ACTCATGACCATCTCGCTT 760
 Db 854 rGlyHisMetGluHisAspHisPheIleVal 864

RESULT 5
 Q870P3 PRELIMINARY; PRT; 906 AA.

AC Q870P3, 24, Created)
 DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 25, Last annotation update)
 DE Hypothetical protein 49D12.100.
 GN 49D12.100.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Hollanć R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BX295540; CAD79685.1; -
 DR InterPro; IPR006598; CAP10.
 DR SMART; SM00672; CAP10; 1.
 DR KW Hypothetical protein.
 SQ SEQUENCE 906 AA; 102137 MW; 71EDE49AF8E03146 CRC64;

Alignment Scores:
 Pred. No.: 0.786 Length: 906
 Score: 89.50 Matches: 60
 Percent Similarity: 39.18% Conservative: 36
 Best Local Similarity: 24.49% Mismatches: 84
 Query Match: 2.34% Indels: 66
 DB: 3 Gaps: 14

US-09-931-733-1 (1-2191) x Q870P3 (1-906)

QY 271 CATGTCTCTGAGTTCGCGCTATAGGTGACGCGAGTTTGTGTATGCTAGTCCCPATTG 330
 Db 550 HisCys---SerMetProProLysArgGluSerAsnProLeuSerCys----- 564
 QY 331 AGACATTACTGCATATAAAATGATTAAAGTAAAGCAGAGAAAATTTCTG----- 382
 Db 565 LysHis-AspProPheLysLeuLeu-----LysArgGlnLeuThrThrLeuAlaTyrGlu 582
 QY 383 -----AAGGTCTATAACTTCCAGGGTCTCTACCTTTGACCTCTGGA 426
 Db 582 uProGluCysLeuThrLysGlyValProPheGln----- 593
 QY 427 TTGGGTTCCTCCAGTGCAGTGCCTTAATATCTCTTCCACTACTGCTTAAAGGCCAATCA 486
 Db 594 -----AsnSerProAlaIleValProSerGlu 603
 QY 487 ATGTATCCAGTAAAGAGTACCCTAAAGCCAGTAAAGATTATCAGCAACACCCAGAGAT 546
 Db 603 uSerArgHisAlaArgSer-----ProValLeuProTyrProProHis----- 617
 QY 547 GTTTATGGCCATATTCAGATCCTTTTGTCT-----TTTCTTGACTTACTCCTTTT 600
 Db 618 -----IleLeuArgGluLeuPheAlaArgAspPheLeuSerAlaCysProArgLys 634
 QY 601 GATAAAGTTATTTCATATGTTT-----CAATCCAGAAATTCAGCTAT 642

```

Db      634 sArgArgMetLeuValTrrpIleLysLysGlyValLeuGlnGlnLeuAspValArgArgLe 654
QY      643 TCAC-----TTCATGGCCCTAGTTTGGATTAAACATCAGCTTCTTTT 687
Db      654 uHisArgLeuGluAspLeuPheProValLeuAlaTyrValSerValIleSerSerIle 674
QY      688 GCTCTGTAGATGGCTTGGACAGTACATATTAACAGTCTATGGCCCTTTCCTACTCATGA 747
Db      674 eSerCysSerLeuSerHisThrIleLysIleProAlaSerPheAlaIleSerSerProPr 694
QY      748 -----CCATCCTCTGCTGCAGACTGCTGCTGGTTAAACACCATCATCTGCTAGTTGA 803
Db      694 oThrProHisProGly-----LysAlaAspGlyArgThrHisAsnPheHisLeuPro-L 712
QY      804 TGCCATTATTATTCACATCTCCTGGGTTTACACGCTAAAGGGCACCAGTGTCTTAAACA 863
Db      712 euPro-----GlnProAsnThrThrSerArgSerArgProSer---LeuArgg 728
QY      864 GAGAAAAATACATGCTGAGCAGAGGATTAAACAGACTCTCTACACTACCGGAAGTTTCAG 923
Db      728 lnProSerLeuLeuSerAlaSerAlaMetProThrSerAlaSerThrProAlaSerThri 748
QY      924 TGCATTTTCCA 934
Db      748 leAsnProPro 751

RESULT 6
Q8HZR7
ID Q8HZR7 PRELIMINARY; PRT; 374 AA.
AC Q8HZR7;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CC chemokine receptor 6.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]_SEQUENCE FROM N.A.
RA Choi Y.K., Fallert B.A., Murphy-Corb M.A., Reinhart T.A.;
RT "Simian Immunodeficiency Virus Dramatically Alters Expression of
RT Homeostatic Chemokines and Dendritic Cell Markers During Infection In
RT Vivo."
RL Blood 0:0-0(2002)
DR EMBL; AF508730; AAN47098.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0016493; F:C-C chemokine receptor activity; IEA.
DR GO; GO:0004872; F:receptor activity; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR004067; CC_6_receptor.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PRO1529; CHEMOKINER6.
DR PRINTS; PRO0237; GPCR_RHODOPSIN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_F1_1; 1.
DR PROSITE; PS0262; G_PROTEIN_RECEP_F1_2; 1.
KW Receptor.
SQ SEQUENCE 374 AA; 42649 MW; A64A515BABC9E928 CRC64;

```

```

Alignment Scores:
Pred. No.: 1.48 Length: 374
Score: 86.50 Matches: 80
Percent Similarity: 34.89% Conservative: 47
Best Local Similarity: 21.98% Mismatches: 153
Query Match: 2.26% Indels: 84
DB: 6 Gaps: 16

```

US-09-931-733-1 (1-2191) x Q8HZR7 (1-374)

```

QY      37 ATTAATATAAATTTAGTCACTGTAATAAATCTAGCTACATAAATATATATATATATATATA 96
Db      51 IleAlaTyrSerLeuIleCysValPheGlyLeuGlyAsnIleLeuValValIleThr 70
QY      97 TTCGCACATTTTGTCTTATCAGGAATTTTGTGTATATTTCTACTAATAGTACTTTTCTGTATG 156
Db      71 PheAlaPheTyrIleLysLysAlaArgSerMetThrAspValTyr---LeuLeuAsnMetAla 89
QY      157 ATAGCAAAATACATTTGTTGGCCACTGCCAAGAGTGGGGGAAAATGCAC----- 208
Db      90 IleAlaAspIleLeuPheVal-LeuThrLeuProPheThrAlaValSerHisAlaThrGln 109
QY      209 -----ATTTTAAACATTTTCTTCTCATCTATGAAAAATTTCTCTTTGGGTATGA 261
Db      109 yAlaTrpValPheSerAsnAlaMetCysLysLeuMetLys-----Gln 123
QY      262 ATTATGAGCATTTGCTCAGTTGCGCCTATAGGTGACCGCAGCATTTGTGTGTATGCTAG 321
Db      123 yIleTyrAlaIleAsnPheAsnCysGlyMet---LeuLeuLeuThrCysIle----- 139
QY      322 TCCCAATTGACACATTTACTGCATATAAAATGATTAATAAGCAGCAAAAATGTTCT 381
Db      140 -----SerMetAspArgTyrIleAlaIleValGlnAlaThrLysSerPheArgLe 156
QY      382 GAAGGTCTATAACTTCCAAGGGTTCTCTACCTTTTGACCTCTGGATTGGGTTCCCGAGT 441
Db      156 uArgTyrArgThrLeuLeuArgSerLysValIleCysLeuIleValTrpGlyGlySerVa 176
QY      442 AGTGCTTAATATCTCTCTCTACTACTGCTTAAGGCCCACTCAATGATGATCCAGCTAA 501
Db      176 lVal-----IleSerSerSerThrPheIlePheAsnGlnLysTyrAsnIleGlnClySe 194
QY      502 GAGTACCTTAAGCCAGTAAAGATTAATCAACAAACACCA-----GAGATGTTAT 552
Db      194 rAspValCysGluProLysTyrGlnThrValSerGluProIleLysTrpLysLeuLeuMe 214
QY      553 GCCCATATTCAGATCTCTTCTTCTTCTTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCT 612
Db      214 lIeuGlyLeuGluLeuLeuPheGlyPhe-----PheIleProLeuMetValIleIlePh 232
QY      613 CATATGGTTT-----CAAATCCAGAAATTCAGTATTCACCTT 648
Db      232 e---TrpTyrMetPheIleValLysThrLeuValGlnAlaGlnAsnSerLysArgHisly 251
QY      649 CATGGCCCTAGTTTGGATTAAACAAATCAGCTTTCTTTTCTCTCTGCTAGTGGCCTTGAC 708
Db      251 sAlaIleArgValIleIleAlaValLeuValPheLeuAlaCysGlnIlePro----- 269
QY      709 AGTACATATTAAACAGTCTATGGCCCTTCCACTCATGACCATCTCTCTGTCAGATCT 768
Db      270 -----HisMetValLeuLeuValThrAl 278
QY      769 GCTG---CTGGGTAAACACCATCATCTGCTAGTGTAGTGAATGCAATTTATTTGACATCTGG 825
Db      278 aAlaAsnLeuGly-AsnMetAsnArgSer-----CysHisSerGluLysLeuLeu- 294
QY      826 GGTITACACGCTAAAGGGCACCAGTGTCTTAAAGCAAAAATACACTGCTGACGA 885
Db      295 -----GlyTyrThrLysThrValThrGluValLeuAlaPheLeuHisC 309
QY      886 GAGGATTAAACAGACTCTCTC----- 905
Db      309 ysCysLeuAsnProValLeuTyrAlaPheIleGlyGlnLysPheArgAsnTyrPheLeuL 329
QY      906 --ACACTACCGGAAGTTCAGTGCATTTTCCAGGATTTTCAAGTCTTCCATGCTGCTCAGT 963
Db      329 ysIleMetLysAspLeuTrpCysVal-ArgArgLysTyrLysSerSerGlyPheSerCys 348
QY      964 TGAGAGTCCATATCCACGTTTGTATTATCCAGGAATTAATGCAGAAAATCTCAGCATTCGT 1023
Db      349 -----AlaGlyArgTyrSerGluAsnIleSerArgGln 359
QY      1024 GAATCT 1029

```

```

Db      360 ThrSer 361
|||||
RESULT 7
O35923
ID O35923 PRELIMINARY; PRT; 3343 AA.
AC O35923;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Breast cancer type 2 susceptibility protein.
GN BCRA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=TESTIS;
RX MEDLINE=97384941; PubMed=9242436;
RA McAllister K.A., Haugen-Strano A., Hagevik S., Brownlee H.A.,
RA Collins N.K., Fucreal P.A., Bennett L.M., Wiseman R.W.;
RL Cancer Res. 57:3121-3125(1997).
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS. ALSO EXPRESSED
CC IN SPLEEN, SKELETAL MUSCLE, THYMUS, MAMMARY GLAND, HEART, OVARY,
CC PROSTATE, LIVER, LUNG, KIDNEY AND BRAIN.
DR EMBL: U89653; AB71378.1; -.
DR PIR: T42207; T42207.
DR InterPro; IPR002093; BCRA2 repeat.
DR Pfam; PF00634; BCRA2; 8.
DR PROSITE; PS50136; BCRA2_REPEAT; 6.
KW Repeat.
FT DOMAIN 991 2050 8 X 25 AA APPROXIMATE REPEATS.
FT REPEAT 991 1015 BCRC1.
FT REPEAT 1204 1228 BCRC2.
FT REPEAT 1412 1436 BCRC3.
FT REPEAT 1510 1534 BCRC4.
FT REPEAT 1645 1669 BCRC5.
FT REPEAT 1828 1845 BCRC6.
FT REPEAT 1946 1970 BCRC7.
FT REPEAT 2026 2050 BCRC8.
SQ SEQUENCE 3343 AA; 372213 MW; 653DB110D2302A8D CRC64;

Alignment Scores:
Pred. No.: 2.32 Length: 3343
Score: 86.50 Matches: 49
Percent Similarity: 38.34% Conservativeness: 25
Best Local Similarity: 25.39% Mismatches: 57
Query Match: 2.27% Indels: 62
DB: 11 Gaps: 9

US-09-931-733-1 (1-2191) x O35923 (1-3343)
QY 1482 AATGATATATAACAAATAAGTTTTTCTTACAAATTCGATGCTGAAGAGTTC 1423
Db |||||
1315 AsnAlaTyrLys-----LeuGluAsnSerAspValSerLysSer 1328
QY 1422 AACATTAGAACTTCTTACCAGCAGACTTACATGTCGATTCATTAGAGACTTATATCCC 1363
Db |||||
1329 ThrSerGlyThr-----ValTyrIleAsn 1336
QY 1362 TTAATGATGCTCAATA-----TAAGAGAAGCAACATGCTTCTTCCCTCCCT 1312
Db |||||
1337 LysGlyAspSerAspLeuProPheAlaGluLysGlyAsnLysTyr-----Pro 1353
QY 1311 TTCCTGTCACAGAGTGGCGATGTCGACGAGAGATCTTGTAGGAGCATCTATTGCTACC 1352
Db |||||
1354 GluSerCysThrGlnTyrValArgGluGluAsnAlaGlnIleLysGluSerValSerAsp 1373
QY 1251 AATACTTTACTGCATCTCCTGGTGGGGAATGCT---TCTATGAATATCGCTATTGTT 1195
Db |||||
1374 LeuThrCysLeuGluValMetLysAlaGluGluThrCysHisMetLysSer----- 1390

```

```

QY 1194 TAAATACTACTGAGCTATTTCATCCAAATTATTAACTGACAAAGCTGGGCTTTCCAGA 1135
Db |||||
1391 -----SerAspLysGluGlnLeu-Prose 1398
QY 1134 TCAGAAAGTCATTATGAATATGTAAATCCTCTCTAAATTTATATGAAGAGTTTAAACAC 1075
Db |||||
1398 rAspLysMetGluGlnAsn-----MetLysGluPheAsnIle 1410
QY 1074 TAGCTGTCTCTTGACCATGACTAGGTACTTTACCTCAATGATTAGATTCACGAATGCT 1015
Db |||||
1410 eSerPhe-----GlnThrAlaSerGlyLysAsnIle 1420
QY 1014 CAGATTTTCTGCATTATTCTCTGGATAAACAAACGCTGGATAGTAGCTCTCAACATGACAG 955
Db |||||
1420 eArgValSerLysGluSerLeuAsnLysSerValAsnIleLeuAspGlnGlu----- 1437
QY 954 CATCGAAGACTTGAAAATCTCTGGAATAACAAACGCTGGATAATGCACATGAAC 918
Db |||||
1438 -ThrGluAspLeuThrValThrSerAspSerLeuAsn 1449

RESULT 8
Q8QGQ5
ID Q8QGQ5 PRELIMINARY; PRT; 409 AA.
AC Q8QGQ5;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-carbamoyl-L-amino acid amidohydrolase.
GN LNC.
OS Geobacillus kaustophilus.
OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1462;
RN [1]
RP SEQUENCE FROM N.A.
RA Chien H.R., Hsu W.;
RT "Cloning, expression, and characterization of the thermostable n-
RT carbamoyl-L-amino acid amidohydrolase from Geobacillus kaustophilus
RT cccr1223."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF425838; AAN31517.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001005; Myb DNA binding.
DR InterPro; IPR002933; Peptidase_M20.
DR Pfam; PF01546; Peptidase_M20; 1.
DR PROSITE; PS00037; MYB_1; 1.
KW Hydrolase.
SQ SEQUENCE 409 AA; 44328 MW; 435771D32179A0B1 CRC64;

Alignment Scores:
Pred. No.: 2.27 Length: 409
Score: 85.00 Matches: 37
Percent Similarity: 36.00% Conservativeness: 17
Best Local Similarity: 24.67% Mismatches: 40
Query Match: 2.23% Indels: 56
DB: 2 Gaps: 6

US-09-931-733-1 (1-2191) x Q8QGQ5 (1-409)
QY 1509 GTACGAAATTACACTTGTATCTC-----ATAAATGTATATAACACAAATAAGTTTTT 1456
Db |||||
262 ValGlyGlnLeuHisValTyrProGlyGlyIleAsnValIleProGluArgValGluPhe 281
QY 1455 TTCTTACAA-----AATTCATAGCT 1435
Db |||||
282 ValLeuAspLeuArgAspLeuLysAlaGluValArgAspGlnValTrpAsnAspIleAla 301
QY 1434 GAAGAGAGTCCCAACATTAGAACTTTCTTACCAGCAGACTTACATGCTAGTCATTAGAG 1375
Db |||||
302 SerArgAlaGluThrIleAlaLysGluArgAsnValArgLeuThrThrGluArgLeuGln 321

```

```

QY 1374 ACTATATCCCTTAATGATGCTCAATATAAGAG-----AAG 1336
Db 322 GluMetAlaProValLeuCySerGluValValLysGlnAlaGluArgAlaCysLys 341
QY 1335 CAACATGGCTCTCTCTCCCTCCCTCTTC-----1309
Db 342 GlnLeuGlyTyr-----ProProPheTrpLeuProSerGlyAlaAlaHisAspGlyVal 359
QY 1308 -----CTGTGC-----1303
Db 360 GlnLeuAlaProLysCysProLysGlyMetIlePheValArgSerGlnAspGlyValSer 379
QY 1302 -----ACAGAGTGGCAGTGGCAGGAGGAGATCTTGTGTAGGAGCATCTATGTCTACC 1252
Db 380 HisSerProAlaGlnTrpSerThrLysGluAspCysAlaValGlyAlaGluValLeuTyr 399
QY 1251 AATACTTTACTGCATCTCTCTGGTTGGGAA 1222
Db 400 HisThrValTrpGlnLeuAlaGlnGlyGlu 409

RESULT 9
Q8PAM4 PRELIMINARY; PRT; 230 AA.
AC Q8PAM4;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein XCC1458.
GN XCC1458.
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Cannavan F., Cardozo J., Chamberg F., Ciapina L.P.,
RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AE012246; AAM40755.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 230 AA; 25843 MW; B8CB70C8DFA911E0 CRC64;

Alignment Scores:
Pred. No.: 2.31 Length: 230
Score: 84.50 Matches: 43
Percent Similarity: 39.30% Conservative: 36
Best Local Similarity: 21.33% Mismatches: 76
Query Match: 46 Indels: 46
DB: 16 Gaps: 9

US-09-931-733-1 (1-2191) x Q8PAM4 (1-230)
QY 603 ATCAAAAAGAGTAAGTCACAGAAAGCAAGAGAGATCTGGAATATGCGCCATAAACATC 544
Db 15 IleArgLysAsnLeuAlaAspHisSerAlaLeuAlaPheArgMetAlaHisGlnIle 34

```

```

QY 543 TCTGGGTGTCGTGATAATAATCTTTTACTGGCTTTAGGGTACTC-----502
Db 35 -----GlyValGlnThrLeuLysAspAlaLysAlaLeu 45
QY 501 -----TTTACGTGGATACATTGAGTGGCTTAAAGCAGTAGGTGAAGAGATATTA 448
Db 46 IleGlyPheSerAspLeuHisAsnIleProLeuLysLeuIleGlyHisAlaGly-----63
QY 447 GGCACCTCACTGGCAACCCCAATCCAGAGT-----CAAAAGGGTAGAGAAC 403
Db 64 -----TrpAlaPheAspLeuLysAsnHisValIleHisProLysArgValHisSer 80
QY 402 CCTTGGAGTTTATGACCTTTCAGAACATTTTTCGCTTTTACTTATATCAATTTTATAT 343
Db 81 LysTrpTrpLeu-----IleLeuProPheIleAlaValAlaValLeuIleValLeuAla 98
QY 342 GCAGTAATGCTCAATTTGGGA-----CTAGCATACACAA 307
Db 99 AlaIleLeuGlyAlaIleAlaAlaAsnAspSerLeuLeuValThrLeuLysGlnThrHis 118
QY 306 ACTGCGGTC---ACCTATAGCGGCAACTCAGACAAATGCTCCATAATTCATACCACAA 250
Db 119 ThrSerIlePheAlaSerGlnThrGluAlaAaGThrGlnThrLeuPheSerValAa3Ser 138
QY 249 GATGAAATTTTCATGATGACGAAGAATAATGCTTAAATAATGTGCAATTTCCCCCACTTT 190
Db 139 AspThrLeuThrValGlnGlnCysAsnThrAlaAsnLysGlnValIle-ProProThrPh 158
QY 189 GTGGCAGTGGCCCAACAATTTGTAATTTGCTATCATACAG-----AAAATCA 142
Db 158 eAspProArgAspLeuLysIleLeuCysGlnValTrpGlnAspProAlaLeuLysSerHi 178
QY 141 TTAGTAGAATATACAAAATTCCTGATAAAGCAAAATGTGCAATATATATATATATATAT 82
Db 178 s-----LeuAlaLysGluValProLysGlnArgAlaAlaSerLeuLeuGlyPheVa 195
QY 81 A 81
Db 195 I 195

RESULT 10
Q9GTJ5 PRELIMINARY; PRT; 507 AA.
AC Q9GTJ5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zinc metalloproteinase 2 MEP2 (Fragment).
OS Ancylostoma caninum (Dog hookworm).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Strongylida;
OC Ancylostomatidae; Ancylostomatidae; Ancylostomatinae; Ancylostoma.
OX NCBI_TaxID=29170;
RN [1]
SEQUENCE FROM N.A.
RA Jones B.F., Hotez P.J.;
RT "Cloning and characterization of a zinc-metalloprotease secreted by
RT the invasive stages of Ancylostoma caninum."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF273705; AAG29105.1; -.
DR HSSP; P08473; 1DMT.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008237; F:metalloproteinase activity; IEA.
DR GO; GO:0004245; F:neprilysin activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000718; Peptidase_M13.
DR InterPro; IPR008753; Peptidase_M13_N.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR Pfam; PF01431; Peptidase_M13; 1.
DR PRINTS; PR00786; NEPRILYSIN_N; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.

```

FT NON_TER 1 1
SQ SEQUENCE 507 AA; 58804 MW; 04B4B185D70F6167 CRC64;

Alignment Scores:

Pred. No.: 3.57 Length: 507
Score: 83.50 Matches: 60
Percent Similarity: 37.00% Conservative: 41
Best Local Similarity: 21.98% Mismatches: 83
Query Match: 2.18% Indels: 89
DB: 5 Gaps: 15

US-09-931-733-1 (1-2191) x Q9H2L2 (1-507)

QY 344 TATATAATGATTAAGTAAAGCCAGAAAATGTTCTGAAGGT-----CATACTTC 397
DB 167 TyrLysAspLeuIleArgSerAlaGlyGlyValMetHisAsnValIleHisAlaPhe 186
QY 398 CAAGGGTCTCTACCTTTCACCTTTCAGCTTGGGTTCCCGAGTGGTCTTAATATCCT 457
DB 187 Gln-----SerMetValAspGlnLeuAspTyrMetSerGlu-----Ala 199
QY 458 TCTTCACCTACTGCTTAAGCCCAACTCAATGATC----- 493
DB 200 ThrLysArgLysAlaIleGluLysSerMetAsnIleIleThrAsnIleAlaPheProAsp 219
QY 494 -----CACGTAAGAGT---ACCTAAAGCCAGTA 520
DB 220 TrpIleMetAspAsnAlaLysLeuAspLeuTyrTyrLysSerIleThrPheAspProThr 239
QY 521 AAGAGTATCACGAACACCCAGAGAGTGTATGCCATATTCAGATCCTCTTTCCTTT 580
DB 240 LysGluAsnTyrTyrAspIleTyrThrLysLeuThrIlePheAsnIleGluAlaGlnTyr 259
QY 581 TCTTGACTTACT-----CCTTTTGTAAAGTATTTCATA--- 616
DB 260 LysHisLeuThrMetAlaThrAlaAspTyrGluGluPheLeuMetProGlyIleVal 279
QY 617 -----TGGTTTCAATCCAGAAATCAGCTATTCATCTGCGCCCTAGTTGATTA 670
DB 280 AsnAlaTyrGlnProGluLeuAsnThrIleThrPheProAlaGlyIle----- 296
QY 671 ACAATCAGCTTCTTTTGTCTGTAGTA-----TGGCCTTGGACA 709
DB 297 -----LeuArgProProTyrPheHisProAspTyrProAlaSer 309
QY 710 GTACATATTAACTCTATGGCCCTTTCACCTCATGACACCATCCTGCTTGCAGATCTG 769
DB 310 IleLysTyrGlyGlyIleGlyLeuIleAlaGlyHisGlu----- 322
QY 770 CTGCTGGTTAAACACCATCATCTGCTAGTTGATGCAATTTATTGACATCCTGGGTT 829
DB 323 -----LeuIle-HisGlyPheAspValGlnGlyVa 332
QY 830 ACAGCAGCT---AAAGGGCACCAGTGTCTTAAACAGAGAAAATACATG----- 878
DB 332 LGlnTyrGlyProLysGlyHis---IleSerTyrProLysLysAsnCysIleGlyTyrP 351
QY 879 -----CTGAGCAGAGATTAAACAGACTCCTCACACTACCGAAGTTCAGTGCAT 928
DB 351 tAspGluGlnSerThrLysGlyPheAsnArgLeu-----AlaGlnCysVa 366
QY 929 TTTCCAGGATTTCAAGCTTCGATGCTGCTGATGTTGAGAGTCCATCCTATCCAGCTTT 988
DB 366 lIleAspGluTyrSerThrPheCysProLeuAspAsnArgThrTyrThrProAsn-Cys 386
QY 989 TA-----TCCAGGAATAATCCAGAAATCTGAGC 1017
DB 386 alaAsnGlyAlaGlnThrGlnGlyGluAsnIleAla 397

RESULT 11

Q9H2L2
ID Q9H2L2 PRELIMINARY; PRT; 333 AA.
AC Q9H2L2;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE GPR18-iso.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Dentritic cell;
RA Xu X., Yang Y., Gao G., Xiao H., Chen Z., Han Z.;
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP261135; AN44671.1; -. membrane; IEA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0001584; F:rhodopsin-like receptor activity; IEA.
DR GO; GO:0007186; P:G-protein coupled receptor protein signalin. .; IEA.
DR InterPro; IPR000276; GPCR_Rhodpsn.
DR Pfam; PF00001; 7tm 1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00262; G PROTEIN RECEPT F1.2; 1.
SQ SEQUENCE 333 AA; 38034 MW; 06B34C9DFF716415 CRC64;

Alignment Scores:

Pred. No.: 3.75 Length: 333
Score: 83.00 Matches: 29
Percent Similarity: 39.52% Conservative: 20
Best Local Similarity: 23.39% Mismatches: 47
Query Match: 2.18% Indels: 28
DB: 4 Gaps: 5

US-09-931-733-1 (1-2191) x Q9H2L2 (1-333)

QY 603 ATCAAAAAGAGTAGTCAAGAAAAGCAAGAGAGTCTGGAAATATGCCATAACATC 544
DB 224 LeuLysProLysValLysGluLysSerIleArgIleIleThrLeuLeuValGlnVal 243
QY 543 TCTGGGTTCGNGATACTTTTACTGGCTTTAGGTTACTCTTTAGCTGGATACATTGA 484
DB 244 LeuValCys-----PheMetProPheHisIleCysPheAlaPheLeu----- 257
QY 483 GTTGGCTTAAGCAGTAGGTGAAGAGGATATTAAAGGCACCTCCTGGGAACCCCAATCC 424
DB 258 -----MetLeuGlyThrGlyGluAsnSerTyr----- 266
QY 423 AGAGGTCAAGGGTAGAGAACCTTGGAGATTATGACCTTCAGAACATTTTCTGCCCTT 364
DB 267 -----AsnProTyrGly-----AlaPheThrThrPheLeuMetAsn 278
QY 363 TTA-----CTATAATCATTTTATATGCAATATGCTCAATGGGACTAGCATAC 313
DB 279 LeuSerThrCysLeuAspValIleLeuTyrTyrIleValSerLysGlnPheGlnAlaArg 298
QY 312 ACACAACTCGGTACCTTACGCGCACTCAGGACAACTCAGGCAATATTCATATACAC 253
DB 299 ValIleSerValMetLeuTyrArgAsnTyrLeuArgSerMetArgLysSerPheArg 318
QY 252 AAAGATCAATTT 241
DB 319 SerGlySerLeu 322

RESULT 12

Q8IB40
ID Q8IB40 PRELIMINARY; PRT; 852 AA.
AC Q8IB40;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE C-13 antigen.
GN MAL8P1.57.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
CX NCBI_TaxID=36329;

QY 1532 AA-----GTCAGAGATTTCAGTGCACATC-----ACCCGAAAAATGTTAA 1573
 Db 642 luGlnIlePheLysGluSerGluTyrLeuIleTyrPheLysAsnAsnIleLeu 662
 QY 1574 CTGTACCCATTAAAGTAATTTCTCATCCGCCATTTCCCTCACCACATGCTCCAGCCCTT 1633
 Db 662 erMetIleLysLysIleLeuPheAspTyrLysIleProGluThrLeuLeuSerAsnPro 682
 QY 1634 GCCAGAACT-----GTTCAATTTCTGCTGCTAACAGAA-----AGA 1672
 Db 682 erGluLysAsnSerLeuIleAsnIleLeuHisLeuGluGluLeuAsnAspPheSerL 702
 QY 1673 AAGCTAAAAAAGCTGCTGGAGAGTCTCCACATTTGTTCTTCTTCTTCTTCTTCTGGG 1732
 Db 702 ysGlnCysHisSerLysHisSerLeuMetGlnTyrPheLysLysIleTyrProG 722
 QY 1733 AATAGCAGAATAGGAGCAAGCCA-----G 1756
 Db 722 InAsnGlnAsnGlnSerAlaGluHisLysLeuArgPheAsnAspAspHisLeuI 742
 QY 1757 CACTAGTCACTAAGTCACTCAACCAAGCCCTTTTCTTCTTCTTCTTCTTCTTCTG 1816
 Db 742 leLysIleSerThrIleHisGlnSerLysGlyLeuGluPhePro----- 756
 QY 1817 ACTTCATTTCTTCTGCTTCTGAGATTAACAACATCTCTCGGTTCCGTTCTTCTGGAAC- 1875
 Db 757 --IleThrTyrLeuProPheIleSerCysTyrThrAsnThrLysArgPheSerAsnAsn 776
 QY 1876 -----T 1876
 Db 776 InLysIleLeuGlnSerThrProTyrAlaAsnLeuGlnSerSerThrLeuAspP 796
 QY 1877 TTACTGATTTATCTCCCTCCACACAATAAGCATTCATCTCTGATTCCTGAGATCT 1936
 Db 796 hePheAspLeuGluArgLeuSerGluAspLeuArgLeuLeu-----TyrValAlaIle 814
 QY 1937 CAAGATCTGGACTACTGTTGAAAAAATTTCCAGGTGAGTACTGTTCTCTGATTTCTTAAAT 1996
 Db 814 hrArgSerIleTyr-----HisCysSerI 822
 QY 1997 ATGATCTTGTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2056
 Db 822 leGlyIleAlaProIleTyrLysSerArgHisLysAsnThrLeuHisThrThrAspL 842
 QY 2057 AGGTCATGGGATGATGGAGTGGGCTGAGTATTTCAGAAAGTCAAACT-----GAGT 2110
 Db 842 euHisHisSerAlaLeuGlyTyrLeuIleGlnAsnLysLysProAsnThrProGlnGlu 862
 QY 2111 TGCTTGATGATGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 2163
 Db 862 euLysGlnCysLeuGlnLysLeuSerIleSerSerAsnLysAsnIleThrPhe 879
 RESULT 14
 Q85483 PRELIMINARY; PRT; 396 AA.
 AC Q85483;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Envelope protein (Fragment).
 OS Rous sarcoma virus.
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.
 ON NCBI_TaxID=11886;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Schmidt-Ruppin subgroup A;
 RX MEDLINE=86263387; PubMed=3014723;
 RA Bova C.A., Manfredi J.P., Swanstrom R.;
 RT "env genes of avian retroviruses: nucleotide sequence and molecular
 RT recombinants define host range determinants.";
 RL Virology 152:343-354 (1996).
 DR EMBL; M14901; AAA87245.1; -
 DR PIR; A47611; A47611.

DR GO; GO:0019031; C:viral envelope; IEA.
 DR InterPro; IPR005166; Avian gp85.
 DR Pfam; PF03708; Avian gp85; 1.
 FT NON_TER 1
 FT CHAIN 57 >396 GLYCOPROTEIN 85.
 FT NON_TER 396 396
 SQ SEQUENCE 396 AA; 43335 MW; 1750744AAIABA261 CRC64;
 Alignment Scores:
 Pred. No.: 4.46 Length: 396
 Score: 82.50 Matches: 58
 Percent Similarity: 36.15% Conservative: 36
 Best Local Similarity: 22.31% Mismatches: 86
 Query Match: 2.18% Indels: 80
 DB: 15 Gaps: 14
 US-09-931-733-1 (1-2191) x Q85483 (1-396)
 QY 1136 CTGGAAAGCCAGGCTTGTCACTTATTATAATTTGGATGATACTTCAGGTAGTATTAA 1195
 Db 61 LeuGluGlnProGly---AsnLeuIrpIleThrIrpAlaAsnArgThrGly----- 76
 QY 1196 ACAATAGCGATTTCATA-----GAAGACATTTTCCCAACACAGGAGATGCGAGT 1243
 Db 77 --GlnThrAspPheCysLeuSerThrGlnSerAlaThrSerProPheGlnThrCys--- 94
 QY 1244 AAGATTTGTAGCAATAGATGCTTACCAAGATCTCTCTGCTGCTGCTGCTGCTGCTG 1303
 Db 95 -----LeuIleGlyIleProSerProIleSerGlu----- 104
 QY 1304 CACAGAAAGGAGGAGGAGAGAGCCATGTTGCTTCTTATATTGAGCATCATTTAAAG 1363
 Db 105 -----GlyAspPheLys 108
 QY 1364 GCAATATAAGTCTCTAATGACTCACAATGCTGCTGGTAAAGATTTCTAAAGTTG 1423
 Db 109 Gly-----TyrValSerAspThrAsnCysThrThrLeuGlyThrAsp----- 122
 QY 1424 GAATCTCTTTCAGCTATCGAATTTTGAAGAAAAAATTTATTTGTTTATATATATTT 1483
 Db 123 ArgLeuValSerSerAlaAspPheThrGlyGlyProAspAsnSerThrThrLeuThrTyr 142
 QY 1484 ATGAGATACAAGTCTAATTTTCGTTACATGATATATGATGATGATGATGATGATGATG 1543
 Db 143 ArgLysValSerCys-----LeuLeuLeuLysLeuAsnVal 154
 QY 1544 TCAGTGTGCATCATCCCGAAAAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 1603
 Db 155 SerMetTrpAspGluProProGluLeu-----GlnLeuLeuGly 167
 QY 1604 CATTTCCCTCACCACATGCTCCAGCCCTTCCAGGAACTGTTTCATTTTCTGACTAA 1663
 Db 168 SerGlnSerLeuProAsnIleThrAsnIleAla----- 178
 QY 1664 CAGAAACGAAAGCTAAAAACACATCGGTGGGAGGAGTCTCCACATTTGTTCTCT 1717
 Db 179 -----GlnIleSerGlyIleThrGlyGlyCysValGlyPheArgProGlnGlyVal 195
 QY 1718 CCATTTTCTTGGGGAATAGCAATAGGAGCAAGC-----CAGCACTATGCTCA 1765
 Db 196 ProTyrTyrLeuGlyTyrSerArgGlnGluAlaThrArgPheLeuLeuArgHisProSer 215
 QY 1766 GCTAACTAAGTACTCAACCAAGCCCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1825
 Db 216 PheSerLysSerThrGluPro-----PheThrValThrAlaAspArg-HisAs 232
 QY 1826 TCTTAGGCTTTCTGAGATTAAC---AACATCTCGCGGTTCGCTTCTCGGAACTTAC 1880
 Db 232 nLeuPheMetGlySerGluTyrCysGlyAlaTyrGlyTyrArgPheThrAsnMetIle 251
 RESULT 15
 Q85483
 ID Q85483 PRELIMINARY; PRT; 446 AA.

Q8HD85;
AC 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE NADH dehydrogenase subunit 4.
GN ND4.
OS Panulirus japonicus (Japanese spiny lobster).
OC Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Palinura;
OC Palinuroidea; Palinuridae; Panulirus.
OX NCBI_TaxID=6736;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2226630; PubMed=12242015;
RA Yamauchi M.M., Miya M.U., Nishida M.
RT "Complete mitochondrial DNA sequence of the Japanese spiny lobster,
RL Panulirus japonicus (Crustacea: Decapoda).";
RL Gene 295:89-96(2002).
DR EMBL; AB071201; BAC16324.1; -.
DR GO; GO:0005739; C:mitochondrion; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR001750; Oxidored q1.
DR InterPro; IPR000260; Oxidored q5_N.
DR Pfam; PF00361; oxidored q1; 1.
DR Pfam; PF01059; oxidored_q5_N; 1.
KW Mitochondrion.
SQ SEQUENCE 446 AA; 49684 MW; 4F4FC11597664EAF CRC64;

Alignment Scores:
Pred. No.: 4.56 Length: 446
Score: 82.50 Matches: 56
Percent Similarity: 37.96% Conservative: 37
Best Local Similarity: 22.86% Mismatches: 77
Query Match: 2.16% Indels: 75
DB: 8 Gaps: 12

US-09-931-733-1 (1-2191) x Q8HD85 (1-446)

QY 173 TTGTTGCCACTGCCAAGTGGGG-----AAATGACATTTTAAAGCATTTCTTT 226
Db 11 LeuMetMetPheAlaGlySerTrpGluValGlnPhePheLeuValValSerPhe 30
QY 227 GCTTCATCTATGAAA-----ATTTTCATCTTTGGTATGAATTATGG 268
Db 31 SerIleSerLeuLysCysAlaHisAspPheMetSerGluLeuGlyTyrMetMetGly 50
QY 269 AGCATTTGCTGATGTCGCCCTATAGTGACCGAGTTTGTGTATGTAGTCCCAAT 328
Db 51 ValAspTyrLeuSerTyrAlaLeuLeuLeuSerAlaTrpValIleAla----- 67
QY 329 TGAGACATTACTGCATATAAATGATTAAAGTAAAGGCGAGAAATGTTCTGAAGGT 388
Db 68 -----LeuMetIleCysSerSerGlnLys---ValMetLysThr 79
QY 389 CATAACTTCCAGGG----- 403
Db 80 TyrAsnPheSerGlyGlyPheIleLeuValAsnMetSerLeuLeuValCysLeuLeu 99
QY 404 ---TTCTTACCTTTCACCTCTGGATTGGGTTCCAGTGGCTTAAATATCCTTCT 460
Db 100 ThrPheSerSerMetAsnTyrLeuPhePhe-----TyrValSer 112
QY 461 TCACCTACTCCTTAGGCCCACTCAATGTATCCACGTAAGAGTACCCCTAAAGCCAGTA 520
Db 113 PheGluSerSerLeuLeuProThrLeuIleLeu-----Ile 124
QY 521 AAAGATTATCAGCAACACCCAGAGATGTTTATGGCC-----ATATTCCAGATC 568
Db 125 LeuGlyTrpGlyTyrGlnProGluArgValGlnAlaGlyLeuTyrMetLeuPheTyrThr 144
QY 569 CTCTTTGCTTTTCTTGACTACTCCTCTTTTGTATAAAGTTATTCATATGTTTCAAAATC 628

Db 145 LeuPheAla-----SerLeuProLeuLeuIleCysLeuPheSerLeuTyrGlnAla 161
QY 629 CAGAAATTCAGCTATT-----CACTTCATGGCCCTA 658
Db 162 GlyGlySerLeuIleLeuGlyLeuProTyrLysValAspGlnMetAspPheIleSerIle 181
QY 659 GTTTGGATTAAACAATCAGCTTTCTTTCTCTGTAGATGGCTTGGACAGTACATATT 718
Db 182 LeuTrpTyrPheSerSerLeuPheAlaPheLeuValLysLeuPro-----LeuTyrLeu 199
QY 719 AACAGTCTATGGGCCCTTTCCACTCAT-----GACCACATCTCTGCTT 760
Db 200 PheHisIleTrpLeuProLysAlaHisValGluAlaProValAlaGlySerMetIleLeu 219
QY 761 GCAGATCTGCTGCTG 775
Db 220 AlaGlyValLeuLeu 224

Search completed: March 28, 2004, 08:12:06
Job time : 192.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 28, 2004, 01:09:44 ; Search time 8590 Seconds
(without alignments)
11055.249 Million cell updates/sec

Title: US-09-931-733-1

Perfect score: 2191

Sequence: 1 tacaataatgagagatttta.....tgaggactgaacaagagaaa 2191

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg.hum.*
31: em.htg.inv.*
32: em.htg.other.*
33: em.htg.mus.*
34: em.htg.pln.*
35: em.htg.rod.*
36: em.htg.mam.*
37: em.htg.vrt.*
38: em.sv.*
39: em.htgo.hum.*
40: em.htgo.mus.*
41: em.htgo.other.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2191	100.0	2191	9	S75415S01	S75415 IF116-inter
2	1997.4	91.2	134222	9	AB045363	AB045363 Hcm sapi
3	1997.4	91.2	300000	9	AP002534	AP002534 Hcm sapi
4	1995.8	91.1	180742	9	AL359753	AL359753 Human DNA
5	1381.8	63.1	9742	6	AX348912	AX348912 Sequence
6	1330.6	60.7	9742	6	AX348911	AX348911 Sequence
7	274.8	12.5	148081	9	AB045361	AB045361 Hcm sapi
8	274.8	12.5	179536	2	AC010814	AC010814 Hcm sapi
9	274.8	12.5	180999	9	AL513205	AL513205 Human DNA
10	240.2	11.0	2775	6	AX188367	AX188367 Sequence
11	240	11.0	2709	9	HUMIFI116A	M63838 Human inter
12	240	11.0	2709	11	G28619	G28619 human STS S
13	240	11.0	4151	9	AF208043	AF208043 Hcm sapi
14	239.4	10.9	504	6	AX884965	AX884965 Sequence
15	239.4	10.9	504	6	BD024575	BD024575 Sequence
16	238.4	10.9	2716	6	AX281850	AX281850 Sequence
17	233	10.6	2709	9	BC017059	BC017059 Homo sapi
18	231.6	10.6	1093	9	HSU30245	U30245 Human myelo
19	230.2	10.5	640	6	AX884966	AX884966 Sequence
20	230.2	10.5	640	6	BD024576	BD024576 Sequence
21	218	9.9	633	9	HUMZE11G11	AF086545 Homo sapi
22	165.2	7.5	15732	6	AX277933	AX277933 Sequence
23	165.2	7.5	15732	6	AX323620	AX323620 Sequence
24	165	7.5	30614	9	AB045362	AB045362 Homo sapi
25	157.4	7.2	15732	6	AX277932	AX277932 Sequence
26	157.4	7.2	15732	6	AX323619	AX323619 Sequence
27	157	7.2	2070	6	AX833326	AX833326 Sequence
28	157	7.2	2070	9	AK094968	AK094968 Homo sapi
29	150	6.8	31816	9	AL513208	AL513208 Human DNA
30	128.4	5.9	141245	2	AC142815	AC142815 Macaca mu
31	122.2	5.6	1597	9	AY185347	AY185347 Homo sapi
32	122.2	5.6	1624	9	AY185346	AY185346 Homo sapi
33	122.2	5.6	1722	9	AY185345	AY185345 Homo sapi
34	122.2	5.6	1749	9	AY185344	AY185344 Homo sapi
35	120.8	5.5	1668	9	BC020822	BC020822 Homo sapi
36	115.6	5.3	459	6	AX884964	AX884964 Sequence
37	115.6	5.3	459	6	BD024574	BD024574 Sequence
38	106.4	4.9	1892	6	AX477979	AX477979 Sequence
39	99.6	4.5	168981	2	AC126347	AC126347 Homo sapi
40	99.6	4.5	206407	9	AL359266	AL359266 Human DNA
41	98	4.5	147894	2	AC021067	AC021067 Homo sapi
42	97.4	4.4	168900	2	AC024418	AC024418 Homo sapi
43	90.4	4.1	146000	9	AP005433	AP005433 Homo sapi
44	90.4	4.1	180940	2	AP000935	AP000935 Homo sapi
45	90.2	4.1	161047	9	AL772182	AL772182 Human DNA

ALIGNMENTS

RESULT 1
S75415S01
LOCUS S75415S01 2191 bp DNA linear PRI 02-JUN-2000
DEFINITION IF116=interferon-inducible myeloid differentiation transcr:ptional
activator [human, Genomic, 2191 nt, segment 1 of 10].
ACCESSION S75415
VERSION S75415.1 GI:913534
KEYWORDS
SEGMENT 1 of 10
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2191)
AUTHORS Trapani,J.A., Dawson,M., Apostolidis,V.A. and Browne,K.A.

Pred. No. is the number of results predicted by chance to have a

Genomic organization of IFI16, an interferon-inducible gene whose expression is associated with human myeloid cell differentiation: correlation of predicted protein domains with exon organization Immunogenetics 40 (6), 415-424 (1994)

95048583

7959953

GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 160709] from the original journal article.

This sequence comes from Fig. 2.

Location/Qualifiers

1..2191

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

source

ORIGIN

```

Query Match      100.0%; Score 2191; DB 9; Length 2191;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2191; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TACATAATTGAGAGATTTTATTCAGAAATATGCTGATTAATATAAATTTAGTCTAGTGA 60
DB 1 TACATAATTGAGAGATTTTATTCAGAAATATGCTGATTAATATAAATTTAGTCTAGTGA 60
QY 61 ATAATGTAGTACATAAATTATATATATATATATATATATTCGCACATTTTTCCTTTATCAGG 120
DB 61 ATAATGTAGTACATAAATTATATATATATATATATATTCGCACATTTTTCCTTTATCAGG 120
QY 121 AATTTTGTATATTTCTACTAATAGATTTTCTGTATGATAGACACAAATTACAATTTGTGC 180
DB 121 AATTTTGTATATTTCTACTAATAGATTTTCTGTATGATAGACACAAATTACAATTTGTGC 180
QY 181 CACTGCCAAGAGTGGGGAAATGCACATTTTAAAGCATTTCTTGTCTCATCTATGNA 240
DB 181 CACTGCCAAGAGTGGGGAAATGCACATTTTAAAGCATTTCTTGTCTCATCTATGNA 240
QY 241 AATTTTCATCTTTTGTGTATGAATTTAGGAGCATTTCTCTGAGTTCGCGCTATAGTGCAC 300
DB 241 AATTTTCATCTTTTGTGTATGAATTTAGGAGCATTTCTCTGAGTTCGCGCTATAGTGCAC 300
QY 301 CGAGTTTGTGTATGCTAGTCCCAATTGAGACATTTACTGCATATAAATGATTAAG 360
DB 301 CGAGTTTGTGTATGCTAGTCCCAATTGAGACATTTACTGCATATAAATGATTAAG 360
QY 361 TAAAGGCGAGAAAATGTTCTCAAGGGTCAATAACTTTCAAGGGTCTCTACCCCTTGACC 420
DB 361 TAAAGGCGAGAAAATGTTCTCAAGGGTCAATAACTTTCAAGGGTCTCTACCCCTTGACC 420
QY 421 TCTGATTTGGGTTCCCAAGTGTGCTTAAATCCTTTTCACTTCTGCTTAAAGGCC 480
DB 421 TCTGATTTGGGTTCCCAAGTGTGCTTAAATCCTTTTCACTTCTGCTTAAAGGCC 480
QY 481 AACTCAATGTATCCAGTAAAGAGTACCCCTAAGCCAGTAAAGATTAATACGACACCC 540
DB 481 AACTCAATGTATCCAGTAAAGAGTACCCCTAAGCCAGTAAAGATTAATACGACACCC 540
QY 541 AGAGATGTTTATGGGCATATTCAGATCCTCTTTTCTTTTCTTCTCACTTACTCTCTTTT 600
DB 541 AGAGATGTTTATGGGCATATTCAGATCCTCTTTTCTTTTCTTCTCACTTACTCTCTTTT 600
QY 601 GATAAAGTTATTCATATGTTTCAATCCAGAAATTCAGTATTTCTATTCATTCATGGCCCTAGT 660
DB 601 GATAAAGTTATTCATATGTTTCAATCCAGAAATTCAGTATTTCTATTCATTCATGGCCCTAGT 660
QY 661 TTGGATTAAACAATCAGCTTTCTTTTCTTGTCTGTAGATGGCCCTGGACAGTACATATTA 720
DB 661 TTGGATTAAACAATCAGCTTTCTTTTCTTGTCTGTAGATGGCCCTGGACAGTACATATTA 720
QY 721 CAGTCTATGGGCCCTTTTCCACTCATGACCAATCCTGCTTGCAGATCTGCTGGGTTA 780
DB 721 CAGTCTATGGGCCCTTTTCCACTCATGACCAATCCTGCTTGCAGATCTGCTGGGTTA 780
QY 781 ACACCACATCATCTGCTAGTTGATGCCATTTATTTGACATCCTGGGGTTACAGCAGCTA 840

```

```

781 ACACCACATCATCTGCTAGTTGATGCCATTTATTTGACATCTCTGGGGTTACAGCAGCTA 840
841 AAGGCGACCCAGTGTCTTAAACAGAGAAAATAACACTGCTGAGCAGAGGATTTAAACAGAC 900
841 AAGGCGACCCAGTGTCTTAAACAGAGAAAATAACACTGCTGAGCAGAGGATTTAAACAGAC 900
901 TCCTCACACTACCGGAAGTTTCAAGTGCATTTTCCAGGATTTTCAAGTCTTCGATGCTCA 960
901 TCCTCACACTACCGGAAGTTTCAAGTGCATTTTCCAGGATTTTCAAGTCTTCGATGCTCA 960
961 TGTGAGAGTCCACTATCCAGGTTTGTATCCAGGATTAATCGAGAAAATTCGAGCAAT 1020
961 TGTGAGAGTCCACTATCCAGGTTTGTATCCAGGATTAATCGAGAAAATTCGAGCAAT 1020
1021 CGTGAATCTAATCATTTGAGGTAAGTAACCTAGTGCATGCTCAAGAGGACAGCTAGTGTTA 1080
1021 CGTGAATCTAATCATTTGAGGTAAGTAACCTAGTGCATGCTCAAGAGGACAGCTAGTGTTA 1080
1081 AACTCCTTCATATAAATTTAAGGAAGGATTTTACATATTTCAATAAGTACTTTCTGATCTGA 1140
1081 AACTCCTTCATATAAATTTAAGGAAGGATTTTACATATTTCAATAAGTACTTTCTGATCTGA 1140
1141 AAGCCCGAGCTTGTGAGTATTAATTAATTTGATGAATACTTTCAAGTACTATTTTAAACAAA 1200
1141 AAGCCCGAGCTTGTGAGTATTAATTAATTTGATGAATACTTTCAAGTACTATTTTAAACAAA 1200
1201 TAGCGATTTCTAAGACATTTTCCCAAACAGGAGATGTCAGTAAAGTATTTGGTAGCAAT 1260
1201 TAGCGATTTCTAAGACATTTTCCCAAACAGGAGATGTCAGTAAAGTATTTGGTAGCAAT 1260
1261 AGATGCTCTTACCAGATCTCTCTGTCACATGTCGTCACAGGAAAGAGGGGA 1320
1261 AGATGCTCTTACCAGATCTCTCTGTCACATGTCGTCACAGGAAAGAGGGGA 1320
1321 GAAGAAGCCATTTGCTTCTTATTTAGAGTCAATTTAAGGGAATATTAAGTCTCTAA 1380
1321 GAAGAAGCCATTTGCTTCTTATTTAGAGTCAATTTAAGGGAATATTAAGTCTCTAA 1380
1381 TGACTCACATGTAAGTGTGGGTAAAGAGTCTTAATGTTGGAATCTCTTTTCAAGTAT 1440
1381 TGACTCACATGTAAGTGTGGGTAAAGAGTCTTAATGTTGGAATCTCTTTTCAAGTAT 1440
1441 CGAATTTTGAAGAAAATAAATCTTATTTGTTTATATATATATATGATGATACAGTAA 1500
1441 CGAATTTTGAAGAAAATAAATCTTATTTGTTTATATATATATATGATGATACAGTAA 1500
1501 TTTCTGTATCATGATATATGATTTGGTGAAGTTCAGAGATTTTCAAGTGTGACATCAACC 1560
1501 TTTCTGTATCATGATATATGATTTGGTGAAGTTCAGAGATTTTCAAGTGTGACATCAACC 1560
1561 CGAAAATGTTTAACTGTACCCATTAAGTAATTTTCTCATCCCCCATTTCCCTTCAACACA 1620
1561 CGAAAATGTTTAACTGTACCCATTAAGTAATTTTCTCATCCCCCATTTCCCTTCAACACA 1620
1621 TGCTCCAGCCCTTCCAGGAACTGTTTCTGCTGACTTAAACAGAAAACGAAAGCTAAA 1680
1621 TGCTCCAGCCCTTCCAGGAACTGTTTCTGCTGACTTAAACAGAAAACGAAAGCTAAA 1680
1681 AACACTGGTGGGAGGAGTCTCCCATTTGTTTCTTACTTCCATTTTCTCTGGGGAATAGCAG 1740
1681 AACACTGGTGGGAGGAGTCTCCCATTTGTTTCTTACTTCCATTTTCTCTGGGGAATAGCAG 1740
1741 AATAGGAGCAGCCAGCAGTACTAGTCACTAAGTCACTCAACCAAGGCCCTTTTTCCT 1800
1741 AATAGGAGCAGCCAGCAGTACTAGTCACTAAGTCACTCAACCAAGGCCCTTTTTCCT 1800
1801 TGTATCTTTGAGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1860
1801 TGTATCTTTGAGATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1860
1861 TCCGTTCTTGGGAATTTTACTGATTTTATCTCCCCCTTCAACAATTAAGCATTTGATTCCT 1920

```


		/note="match IFI16 (NM_005531) with account of intervene sequences."	
STS		complement(241289..241444) /notes="R448430;The location is between each flanking site of PCR primers."	
STS		complement(272376..272547) /notes="R434584;The location is between each flanking site of PCR primers." /db_xref="GDB:355027" 277112..277257 /gene="IFI16"	
STS		/note="RH103;The location is between each flanking site of PCR primers." /db_xref="GDB:4559702" 277120..277321 /genes="IFI16"	
STS		/note="RH37786;The location is between each flanking site of PCR primers." /db_xref="GDB:6455398" 282987..283075 /notes="SHGC-18224;The location is between each flanking site of PCR primers." /db_xref="GDB:1230561" complement(284696..289063) /gene="AIM2"	
gene		complement(284696..289063) /genes="AIM2"	
misc_feature		/note="match AIM2 (NM_004833) with account of intervene sequences."	
ORIGIN			
	Query Match	91.2%; Score 1997.4; DB 9; Length 300000;	
	Best Local Similarity	98.4%; Pred.No. 0;	
	Matches 2178; Conservative	0; Mismatches 11; Indels 24; Gaps 15;	
QY	1	TACATAATTTCAGAGATTTTATTCAGAAATATCGCTGAATTAATATAAATTAGTCACTGT	60
Db	230415	TACATAAATTGAGAGATTTTATTCAGAAATATCGCTGAATTAATATAAATTTAGTCACTGT	230474
QY	61	ATAATAGTACTACATAA-TTATATATATGTATATATATTCGGCACA-TTTTTGCTTTATCA	118
Db	230475	ATAATAGTACTACATAA-TTATATATATGTATATATATTCGGCACA-TTTTTGCTTTATCA	230534
QY	119	GGAAATTTTGATATTTCTACTAATAGCTTTCTGTATGATAGACAAAATACAAATGTTG	178
Db	230535	GGAAATTTTGATATTTCTACTAATAGCTTTCTGTATGATAGACAAAATACAAATGTTG	230594
QY	179	GCCACTGCCAACAGTGGGGAAAAATGCACATTTTAAAGCATTTCTTGGCTTCATCTATG	238
Db	230595	GCCACTGCCAACAGTGGGGAAAAATGCACATTTTAAAGCATTTCTTGGCTTCATCTATG	230654
QY	239	AAAAATTCATCTTTGTGTPATGAATATAGGACATTTGCTGAGTTGGCGCTATAGGTG	298
Db	230655	AAAAATTCATCTTTGTGTPATGAATATAGGACATTTGCTGAGTTGGCGCTATAGGTG	230713
QY	299	ACGCAGATTTGTGTPATGCT-----AGTCCCAATGA-GACATPACTGCATATAAAATG	352
Db	230714	ACGCAGATTTGTGTPATGCTGGAGTCCCACATTTGAGGACATTTACTGCATATAAAATG	230773
QY	353	ATTATAAGTAAAGGCAGAAAATCTCTGAGGGTCAATACTCCAAGGGTTCTCTACC	412
Db	230774	ATTATAAGTAAAGGCAG-AAATGTTCTGAAGGGTCAATACTCCAAGGGTTCTCTACC	230832
QY	413	CTTTGACCTCTGGATTTGGGGTTCCCGAGTGAAGTCCCTTAATATCTCTTCTACCTACTGCC	472
Db	230833	CTTTGACCTCTGGATTTGGGGTTCCCGAGTGAAGTCCCTTAATATCTCTTCTACCTACTGCC	230892
QY	473	TTAAGGCCAACTCAATGTATCCAGTAAAGAGTACCCCTAAAGCCAGTAAAGATTAATCAC	532
Db	230893	TTAAGGCCAACTCAATGTATCCAGTAAAGAGTACCCCTAAAGCCAGTAAAGATTAATCAC	230952
QY	533	GAACACCCAGAGATGTTTATGCCAAT-ATTCCAGATCTCTTCTGTTTCTTGACATTAC	591

QY	712	ACATATTACAGTCTATGGG---CCCTTTCCACTCATGACCACCATCTCTGCTGCGATCT	768	QY	1787	AGGCTTTTCTCTGTTATCTTTGCGATACTTCAATTTCTTAGGTTCTGGAATTA	1846
Db	105763	ACATATTACAGTCTATGGGCGCCCTTTCCACTCATGACCACCATCTCTGCTGCGATCT	105822	Db	106843	AGGCTTTTCTCTGTTATCTTTGCGATACTTCAATTTCTTAGGTTCTGGAATTA	106902
QY	769	GCTGCTGGGTTACACACCATCATCTGCTAGTTGATGCCATTTATTTGACATCTCTGGGT	828	QY	1847	CAACATCTCTGCGGTTCCGTTTCTGGAACTTTTACTGATTTACTCTCCCTCACACAAT	1906
Db	105823	GCTGCTGGGTTACACACCATCATCTGCTAGTTGATGCCATTTATTTGACATCTCTGGGT	105882	Db	106903	CAACATCTCTGCGGTTCCGTTTCTGGAACTTTTACTGATTTACTCTCCCTCACACAAT	106962
QY	829	TACAGCACGCTAAAGGSCACCCAGTGTCTTAAACAGAGAAAAATACACTCTGAGCAGAG	888	QY	1907	AGCATTGATTCCTGATCTTCTGAAGATCTCAAGATCTGAGTCTGTTGAAAAATTTTC	1966
Db	105883	TACAGCACGCTAAAGGSCACCCAGTGTCTTAAACAGAGAAAAATACACTCTGAGCAGAG	105942	Db	106963	AGCATTGATTCCTGATCTTCTGAAGATCTCAAGATCTGAGTCTGTTGAAAAATTTTC	107022
QY	889	GATTAAACAGACTCTCTACACTACCGGAAGTTTCAAGTCAATTTCCAGGATTTTCAAGTCT	948	QY	1967	C-----AGGTGAGTACTGTTCTCTGATTTCTTAATATGATCTTCTCTCTGAAAT-C	2021
Db	105943	GATTAAACAGACTCTCTACACTACCGGAAGTTTCAAGTCAATTTCCAGGATTTTCAAGTCT	106002	Db	107023	CAGTGGGTGAGTACTGTTCTCTGATTTTCTTAATATGATCTTCTCTCTGAAATTC	107082
QY	949	TCCATGCTGTCTGCTGAGAGTCCACTATCCAGTTTGTGTTTATCCAGGAATAATGCAGAA	1008	QY	2022	CCAGATCAACAGGGGACAATCAGTATT-GTTATTCAGGCTCATGGATGATGGGAATAG	2080
Db	106003	TCCATGCTGTCTGCTGAGAGTCCACTATCCAGTTTGTGTTTATCCAGGAATAATGCAGAA	106062	Db	107083	CCAGATCAACAGGGGACAATCAGTATTGTTTATTCAGGCTCATGGATGATGGGAATAG	107142
QY	1009	AATCTGAGCAATCTGTAATCTAATCAATGAGGTAAGTAACTGATGATGATGATGATGATG	1068	QY	2081	GGCTGAGTATTCAGAAAGTG-AACTGAGTTGCTTGATATGATGATGATGATGATGATG	2139
Db	106063	AATCTGAGCAATCTGTAATCTAATCAATGAGGTAAGTAACTGATGATGATGATGATGATG	106122	Db	107143	GGCTGAGTATTCAGAAAGTGAAACTGAGTTGCTTGATATGATGATGATGATGATGATG	107202
QY	1069	CAGCTAGTGTAACTCTTATATATAATTAAGGAGGATTTTACATATTCAATGACT	1128	QY	2140	GAAGATAACAGGCACTCTTCTATT-CACCACACTGAGGACTGAACAGAGAAA	2191
Db	106123	CAGCTAGTGTAACTCTTATATATAATTAAGGAGGATTTTACATATTCAATGACT	106182	Db	107203	GAAGATAACAGGCACTCTTCTATTCCACCACACTGAGGACTGAACAGAGAAA	107255
QY	1129	TTCTGATCTCGAAAGCCAGGCTGTGCTGATTAATAATTAATGATGATGATGATGATGATG	1188	RESULT 5			
Db	106183	TTCTGATCTCGAAAGCCAGGCTGTGCTGATTAATAATTAATGATGATGATGATGATGATG	106242	AX348912/c			
QY	1189	TATTTAAACAAATAGCGATTTCTATGAGGTAAGTAACTGATGATGATGATGATGATGATG	1248	LOCUS	AX348912	9742 bp	DNA linear PAT 06-FEB-2002
Db	106243	TATTTAAACAAATAGCGATTTCTATGAGGTAAGTAACTGATGATGATGATGATGATGATG	106302	DEFINITION	Sequence 370 from Patent WO0202807.		
QY	1249	ATTGTTAGCAATAGATGCTCTACCAAGATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG	1308	ACCESSION	AX348912		
Db	106303	ATTGTTAGCAATAGATGCTCTACCAAGATCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG	106362	VERSION	AX348912.1	GI:18614947	
QY	1309	GAAGAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	1368	KEYWORDS	synthetic construct		
Db	106363	GAAGAGGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	106422	SOURCE	synthetic construct		
QY	1369	ATAAGTCTTAATGACTCACATGTAAGTCTGCTGGGTAAGAAAGTTCTAATGTTGAACT	1428	ORGANISM	artificial sequences.		
Db	106423	ATAAGTCTTAATGACTCACATGTAAGTCTGCTGGGTAAGAAAGTTCTAATGTTGAACT	106482	REFERENCE	1		
QY	1429	CCTTTGAGCTATCGAATTTTCTAAG-AAAAAACTTATTTGTTTATATACATTTATGA	1487	AUTHORS	Olek, A., Piepenbrock, C. and Berlin, K.		
Db	106483	CCTTTGAGCTATCGAATTTTCTAAG-AAAAAACTTATTTGTTTATATACATTTATGA	106542	TITLE	Diagnosis of diseases associated with cell signalling		
QY	1488	GATACAAAGTGTAAATTCGTTACATGATATATGCTATTTGTTGAGTCAAGATTTTCAAG	1547	JOURNAL	Patent: WO 0202807-A 370 10-JAN-2002;		
Db	106543	GATACAAAGTGTAAATTCGTTACATGATATATGCTATTTGTTGAGTCAAGATTTTCAAG	106602	EPIGENOMICS	AG (DE)		
QY	1548	TGTGCACATACCCGAAATGTTAACTGTACCCATTAAGTAATTTCTCATCCCCCATTT	1607	FEATURES	Location/Qualifiers		
Db	106603	TGTGCACATACCCGAAATGTTAACTGTACCCATTAAGTAATTTCTCATCCCCCATTT	106662	source	1. .9742		
QY	1608	TCCCTTCACCATGCTCCAGCCCTTCCAGGAACTGTTTCATTTTCTGACTTAACAGA	1667	organism	/organism="synthetic construct"		
Db	106663	TCCCTTCACCATGCTCCAGCCCTTCCAGGAACTGTTTCATTTTCTGACTTAACAGA	106722	/mol_type	/mol_type="unassigned DNA"		
QY	1668	AACGAAAGCTAAACAACTGTTGGGAGGAGTCTCCACATTTTCTTCTACTCCATTTTCTC	1727	/db_xref	/db_xref="taxon:32630"		
Db	106723	AACGAAAGCTAAACAACTGTTGGGAGGAGTCTCCACATTTTCTTCTACTCCATTTTCTC	106782	/note	/note="chemically treated genomic DNA (Homo sapiens)"		
QY	1728	TGGGG-ANTAGCAGAAATAGGAGCAGCAGCAGTACTGATGATGATGATGATGATGATGATG	1786	Query Match	63.1%; Score 1381.8; DB 6; Length 9742;		
Db	106783	TGGGGCAATAGCAGAAATAGGAGCAGCAGCAGTACTGATGATGATGATGATGATGATGATG	106842	Best Local Similarity	80.8%; Pred. No. 2.7e-269;		
				Matches 1787; Conservative 0; Mismatches 402; Indels 24; Gaps 14;			

6242 AAAATTTCACTTTTATATATAAATTAATAAAGATTATCTTAATATACC-CTTAATAATA 6184
299 ACCGACGTTGTGTGTATGCTAGT-CCCAATTGAGACATTTACTGCAATATAAATAAG 352
6183 ACCGCAATTTATATATATATATATAAATCCCAATTAATAAACAATTACTACATATAAATA 6124
353 ATTATATAGTAAAGGCGAGAAATGTTCTGAGGCTCATTAATCTCCAAAGGTTCTCTACC 412
6123 ATTATAAATTAATAAACA-AAAAATTTCTAAATAATCATAACTTCCAAAATTTCTTACC 6065
413 CTTTGACCTCTGAGTTGGGTTCCAGTAGTGCCTTAATATCTCTTCACTACTGACC 472
6064 CTTTAACTCTAATTAATAATTTCCCAATTAATCTTAAATATCTTCTTACTTACC 6005
473 TTAAGGCCAATCTAATGTATCCAGTAAAGAGTACCTTAAGCCAGGTAAGATATAC 532
6004 TTAAGCCAACTCAATATATCCAGTAAATAATACCTTAAGCCAAATAAATAATATAC 5945
533 GAACACCCAGAGATGTTTATGGCCAT-ATTCCAGATCTCTTCTTCTTCTTCTGACTTAC 591
5944 GAACACCCAAATAATTTATAACCATCATTTCCAAATCTCTTACTTTTCTTAACTTAC 5885
592 TCTTTTTTTCATAAAGTTATTCATATGTTTCAATCCAGAAATTCAGTATTTCACTTTCAT 651
5884 TCTTTTTTAAATAAATTTATATATATTTCAATCCAAATTCAACTTTCATTTCTTCT 5825
652 GGCCCTAGTTTGGATTAATAACATGAGTTTCTTCTGCTGTAGATGGGCTTGGACGT 711
5824 AACCTTAATTTAAATTAATAAACAATCAACTTTCTTTTACTCTATTAATAAACCCTTAAACAAT 5765
712 ACATATTAACAGTCTAT- -GGGCCCTTTCACATCATGACCATCTGCTTGTGAGATCT 768
5764 ACATTTAAACATCTATAAACCCCTTTTCCACTATAACCATCTCTTACCAATCT 5705
769 GCTGCTGGGTTAAACACACATCATCTGCTAGTTGATGCCATTTATTTGACATCTCTGGGT 828
5704 ACTACTAAATTAACACACATCATCTACTAATTAATACCATTTATTTAAACATCTTAAAT 5645
829 TAGACGAGCTTAAGGCGACGAGTGTCTTTAAGAGAAATAATACACTGCTGAGCGAG 888
5644 TACAACACGCTTAATAAACAACCCCACTATCTTTAACAATAAATAAATAAATAAATAA 5585
889 GATTAAACAGACTCCTCACACTACCGAAGTTCAGTGCATTTTCCAGGATTTTCAAGTCT 948
5584 AATTAAACAACTCCTCACACTACCGAATTTCAATACATTTTCCAAAATTTTCAATCT 5525
949 TCGATGCTGCTAGTTGAGAGTCCACTATCCAGTTTGTATCCAGGAATAATGCGAA 1008
5524 TCGATACTATCATATTAATAAATCCACTATCCAGTTTATTTATCCAAAATAATACAAA 5465
1009 AATCTGAGCATTCGTGAATCTAATCATGAGGTAAAGTAACTGATGATGCTCAAGGGA 1068
5464 AATCTAAACATTCGTAAATCTAATCATATTAATAAATAAATAAATAAATAAATAAATA 5405
1069 CAGCTAGTGTAAATCTCTTATATATAAATTTAAGGAGGATTTACATATTTCAATAATGACT 1128
5404 CACTAATAATTAACCTCTTATATAAATTTAATAAATAAATAAATAAATAAATAAATA 5345
1129 TCTGATCTGGAAAGCCAGGCTGTGTCAGTTATTAATAATTTGGATGAATATCTTCAGTAG 1188
5344 TTCTAATCTAAAAAACCACAACTTATCAATTAATAATAATAAATAAATAAATAAATAA 5285
1189 TATTAAACAAATAGGATTTCTAGAGACATTTCCCAACGAGGATGCGAGTAAAGT 1248
5284 TATTAAACAAATAGGATTTCTAATAAACAATTTCCCAACGAGGATGCGAGTAAAGT 5225
1249 ATTGGTAGCAATAGATGCTCTCAACAGATCTCTCTGCTGCACTGCGACCTCTGCGACAG 1308
5224 ATTAATAACAAATTAATCTCTCAACAAATCTCTCTTACCATCTACCATCTTATACAA 5165
1309 GAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1368

5164 AAAAAAAAAAAAAAAAAAACCATATTAATTTCTTCTTCTTATATTTTAAACATCATTTTAAAAAAT 5105
1369 ATAAGTCTCTAATGACTCACATGTAAGTCTGCTGGGTAAGAAAGTTCTTAATGTTGAACT 1428
5104 ATAAATCTCTAATAAATCTCACATATAATCTACTAAATAAATAAATAAATCTTAATATTAAC 5045
1429 CTTTTCAGCTATCGAATTTTGT- AAGGAAATAAATAAATAAATAAATAAATAAATAAATAA 1487
5044 CTTTTCAGCTATCGAATTTTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 4985
1488 GATACAGTCTAATTTCTGTTACATGATATATGCAATTTGGTCAAGTCAAGATTTTCAG 1547
4984 AATACAAATATAATTTGTTTACATAAATAATATACCATATTAATAAATAAATAAATAA 4925
1548 TGTGCACATCACCGGAAATAATGTTTACCTGACCAATTAAGATTAATTTCTCATCCCCCAT 1607
4924 TATACATCACCGGAAATAATTAACATAACCACTTAAATAAATAAATAAATAAATAAATA 4865
1608 TCCCTCACACATGCTCCAGCCCTTCCAGGAACTGCTCCCAATTTGTTCTCTGACTATCAGA 1667
4864 TCCCTCACACATGCTCCAGCCCTTCCAGGAACTGCTCCCAATTTGTTCTCTGACTATCAGA 4805
1668 AAGGAAAGCTTAAATAAACAATGTTGGAGGAGTCTCCCAATTTGTTCTCTGACTATCAGA 1727
4804 AAGGAAAGCTTAAATAAACAATGTTGGAGGAGTCTCCCAATTTGTTCTCTGACTATCAGA 4745
1728 T-GGGGAATAGCAGATATAGGAGCAAGCAGCACTAGTCACTAACTAAGTGAATCAACCA 1786
4744 TAAAACAATTAACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4685
1787 AGGCTTTTCTGTTATCTTTGAGAGTACTTCAATTTCTTGGGTTCTGAGGATTA 1846
4684 AAGCTTTTCTGTTATCTTTTACAAATCTTCAATTTCTTAAAGTTCTTAAATAAATA 4625
1847 CAACATCTGCGGTTCCGTTTCTGGGAATTTTACTGATTTATCTCCCTCCCTCACAAAT 1906
4624 CAACATCTGAGTCCGTTTCTTAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4565
1907 AAGCATTGATCTGCAATTTCTGAGATCTCAAGATCTGAGACTGCTGTTGAAAAATTTTC 1966
4564 AAGCATTGATCTGCAATTTCTGAGATCTCAAGATCTGAGACTGCTGTTGAAAAATTTTC 4505
1967 C- - -AGGTGAGTACTGTTCTGATTTTGTAAATAATGATCTGTTCTCTTCTGAGT-C 2021
4504 CATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 4445
2022 CCAGAAATCAAGGGGACAACTCAGTATTT- GTTATTCAGGCTCATGGGATGATGGGAGTAG 2080
4444 CCAGAAATCAAGGGGACAACTCAGTATTTTAAATAAATAAATAAATAAATAAATAAATA 4385
2081 GGCTGAGTATTCAGAAAGT- GAAACTGAGTTCCTTGTATGATGATCTCTTCAATTTACTTAG 2139
4384 AACTAAATATTCAGAAAGTTCCTTGTATGATGATGATCTCTTCAATTTACTTAGTAA 4325
2140 GAAGATAACAGGCTCTTCTTATTT- CACCACAACTGAGGACTGAAAGAGATA 2191
4324 AATAATACAAATCTTCTTATTTCCACCACTAATAAATAAATAAATAAATAAATAAATA 4272

RESULT 6

AX348911

LOCUS

DEFINITION

AX348911

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

Sequence 369 from Patent WO0202807.

AX348911.1 GI:18614946

synthetic construct

synthetic construct

artificial sequences.

1

Olek, A., Piepenbrock, C. and Berlin, K.

Diagnosis of diseases associated with cell signalling

Patent: WO 0202807-A 369 10-JAN-2002;

AX348911

9742 bp

DNA

linear

PAT 06-FEB-2002

FEATURES		Epigenomics AG (DE)		Location/Qualifiers		Source					
		1..9742		/organism="synthetic construct"		/mol_type="unassigned DNA"					
				/db_xref="taxon:32630"		/note="Chemically treated genomic DNA (Homo sapiens)"					
ORIGIN											
Query Match		60.7%; Score 1330.6; DB 6; Length 9742;									
Best Local Similarity		79.3%; Pred. No. 6e-259;									
Matches 1755; Conservative		0; Mismatches 434; Indels 24; Gaps 14;									
QY	1	TACATAATTGAGAGATTTTATTTCAGAAATATCCCTGATTAATATAAAATTTAGTCACTGTA	60								
Db	3261	TATATAAATTGAGAGATTTTATTTCAGAAATATCCCTGATTAATATAAAATTTAGTCACTGTA	3320								
QY	61	ATAAATGTAGTACATAAA-TTATATATATGTATATATATATTCGCACA-TTTTGTCTTTATCA	118								
Db	3321	ATAAATGTAGTTATATAAATTATATATATGTATATATATTCGTATATTTTGTGTTTATTA	3380								
QY	119	GGAAATTTTGTATATTCTACTAATGACTTTTCTGTATGATAGACACAAATTACAATTTGTTG	178								
Db	3381	GGAAATTTTGTATATTTTATTAATGATTTTTTTTGTATGATAGTATAAAATTAATAATTTGTTG	3440								
QY	179	GCCACTGCCACAAGTGGGGGAAATGACATTTTTTAAAGCAATTTCTTTGCTTCATCTATG	238								
Db	3441	GTTATTGTTTATAAAGTGGGGGAAATGATATATTTTAAAGTATTTTTTTGTTTATTTATG	3500								
QY	239	AAAAATTCATCTTTGTGGTATGAAATATGGAGCATTTGCTCGAGTTGCGCGCTATAGGTTG	298								
Db	3501	AAAAATTTTATTTTGTGGTATGAAATATGGAGTATTTGTTTGGTTG-TTTTTATAGGTTG	3559								
QY	299	ACGCGAGTTTGTGTG-T-----ATCCTAGTCCCAATTTGAGACATTTACTGCATATAAAATG	352								
Db	3560	ATCGTAGTTTGTGTGTTGATGTTGGAGTTTAAATTTGAGGATATTTATGATATAAAATG	3619								
QY	353	ATTATAAGTAAAGGAGAAAATGTTCTGAAGGTGATATACTTTCCAAGGGTTCTCTACC	412								
Db	3620	ATTATAAGTAAAGGATAG-AAAAATGTTTGAAGGGTTATAATTTTGAAGGGTTTTTATT	3678								
QY	413	CTTTGACCTCTGATTTGGGGTTCCAGTCAGTGCCTTAATATCTCTCTCACTACTGCC	472								
Db	3679	TTTTGATTTTGGATTGGGGTTTTTGTAGTGATTTTAAATATTTTTTTTATTTATTTGTT	3738								
QY	473	TTAAGGCCAACTCAATGTATCCACGTAAAGAGTACCTTAAAGCCAGTAAAGATATCAC	532								
Db	3739	TTAAGGTTAAATTAATGTATTTACGTAAAGAGTATTTTAAAGTTAGTAAAGATTTATAC	3798								
QY	533	GAACACCCAGAGATGTTTATGGCCA-TATTCAGATCCCTCTTTGCTTTTCTTGACTTAC	591								
Db	3799	GAATATTTAGAGATGTTTATGGTTATTTATTTAGATTTTTTTTGTGTTTTTTTGATTTAT	3858								
QY	592	TCCTTTTGTATAAAGTTATTCATATGGTTTCAAAATCCAGAATTCAGCTATTTCACTTCAT	651								
Db	3859	TTTTTTTGTATAAAGTTATTTATATATGTTTAAATTTAGAAATTTAGTTATTTATTTTAT	3918								
QY	652	GGCCCTAGTTTGGATTAAACAATCAGCTTTCTTTTGTCTGTAGATGGCCCTTGACAGT	711								
Db	3919	GGTTTGTAGTTTGGATTAAATTAATTTAGTTTTTTTTTGTGTTGATGTTTTGGATAGT	3978								
QY	712	ACATATTAACAGTCTATGGG---CCCTTTTCCACTCATGACCACATCCTCTGCTTGAGATCT	768								
Db	3979	ATATATTAATAGTTTATGGGGTTTTTTTTTTTATTTATGATTAATTTTGTGTTGTAGATTT	4038								
QY	769	GCTGCTGGGTTAACACCATCATCTGCTAGTTGATGCCATTTATTTGACATCCTGGGGT	828								
Db	4039	GTTGTTGGGTTAATATTAATTTATTTTGTAGTTGATGTTTATTTATTTGATATTTTGGGGT	4098								
QY	829	TACAGCACGCTAAAGGGCCACCGAGTCTCCTTAACAGAGAAAATACACCTGCTGAGCAGAG	888								
Db	4099	TATAGTACGTTAAAGGGTATTTATTTGTTTTTAAATAGAGAAAATATATTTGTTGAGTAGAG	4158								
QY	889	GATTAAACAGACTCCTCACACTACCGAAGTTTCAGTGCATTTTCCAGATTTTTCAGTCT	948								
Db	4159	GATTAAATAGATTTTTTTTATATATATTCGGAAGTTTATGTTATTTTATGAGTTTTTAAAGTTT	4218								
QY	949	TCGATGCTGTCACTGTTCAGAGTCCACTATCCACCTTTGTTTATCCAGGAATAATGAGAA	1008								
Db	4219	TCGATGTTGTATGTTGAGAGTTTATTTATTCGTTTGTGTTTATTTAGGAATAATGAGAA	4278								
QY	1009	AATCTGAGCATTCGTGAATCTAATCATAGGTAAGTAACCTAGTCACTGTCAGAGGA	1068								
Db	4279	AATTTGAGTATTCGTGAATTTAATTTAGGTAAGTAATTTATGTTTATGTTTAAAGGGA	4338								
QY	1069	CAGCTAGTGTAAACTCCTTCATATAAATTTTAAAGGAAGATTTACATATTTCAATAACACT	1128								
Db	4339	TAGTTAGTGTAAATTTTATATATAAATTTTAAAGGAAGATTTATATAATTTATATAACACT	4398								
QY	1129	TTCTGATCTGAAAGCCAGGCTTTGTGAGTTTATTAATTAATTTGGAATTAATCTTCAGGTAG	1188								
Db	4399	TTTTGATTTGGAAGTTTGTAGTTTATTAATTTAAATTTGGAATTAATTTTGTAGGTAG	4458								
QY	1189	TATTTAAACAAATAGCGAATTTCTATAGAGACATTTCCCAACAGGAGATCGAGTAAAGT	1248								
Db	4459	TATTTAAATAAATAGCGAATTTTATAGTAGATATTTTTTTTTAAATTTAGGAGATGCTGTAAGT	4518								
QY	1249	ATTGCTAGCAATAGATGCTCCTACCAAGATCTCCTGCCACTGCCACTCTGTGCACAG	1308								
Db	4519	ATTGCTAGTAATAGATGCTTTTATTAAGATTTTTTTTTGTTATTTGTTATTTTGTGTATAG	4578								
QY	1309	GAAGGAGGGGAGAGAACCATGTTGCTCTCTTATATTTTGGAGCATCATTTAAGAGGAAT	1368								
Db	4579	GAAGGAGGGGAGAGAAAGTTATGTTGTTTTTATATTTTATTTTATTTTAAAGAGGAAT	4638								
QY	1369	ATAAGTCTCTAATGACTCACATGTAAGTCTGCTGGGTAAAGAAAGTTCTAATGTTGTAAGT	1428								
Db	4639	ATAAGTTTTTAAATGATTTATATGTAAGTTTCTGGTAAGAAAGTTTAAATGTTGTAAGT	4698								
QY	1429	CTTTTCAGCTATCGAATTTTGTGTAAG-AAAAAACCTTATTTGCTGTTTATATATCATTTATGA	1487								
Db	4699	TTTTTTTATGTAATCGAATTTTGTAAAGAAAATAATTTATGTTTATATAATTTATGA	4758								
QY	1488	GATACAACTGTAATTTCTGTTACATGATATATGCCATATTTGGTGAAGTCAAGATTTTCAG	1547								
Db	4759	GATATAAGTGTAAATTTCTGTATATGATATATTTCTGTGAAGTTAGAGATTTTATAG	4818								
QY	1548	TGTGCACATCACCAGAAAATGTTAACTGTACCCATTAAGTAATTTCTCATCCCTCAAT	1607								
Db	4819	TGTGTATATTAATTCGAAAAATGTTAAATTTGTAATTTTAAAGTAATTTTATTTTATTT	4878								
QY	1608	TCCCTCACCACATGCTCCAGCCCTTGCAGAAACTGTTCTATTTTCTCTGACTAACAGA	1667								
Db	4879	TTTTTTTATATATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTT	4938								
QY	1668	AACGAAAGCTAAAAACACTGTTGGGAGAGTCTCCACATTTGTTTCTTCTCACTCCCTC	1727								
Db	4939	AACGAAAGCTAAAAATAATTTGGGAGAGTCTTTTATATGTTTATGTTTATGTTTATGTTT	4998								
QY	1728	TGGGG-ATATAGCAGATAGGAGCAGCCAGCAGTACTAGTCACTAAGTCACTCAACCA	1786								
Db	4999	TGGGGTAATAGTAATAAGGATAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTTAGTT	5058								
QY	1787	AGGCTTTTTTCTGTTTATCTTTTGCAGATACTTTCTATTTTCTTAGCGTTTCTGGAGATTA	1846								
Db	5059	AGGTTTTTTTTTGTGTTTATTTTGTAGATATTTTATTTTGTAGCGTTTGTGGAGATTA	5118								
QY	1847	CAACATCTCGGGTTCGTTTCTGGGAACCTTTACTGATTTATCTCCCGCTTCCACACAAAT	1906								
Db	5119	TAAATTTTGGGTTTCTGTTTGGGAATTTTATGATTTTATTTTATTTTATTTATTAATAAT	5178								
QY	1907	AAGCATTCATTTCTGCTGCTTCTGGAATCTTCAAGATCTCGGACTACTGTTTGAAGAAATTTT	1965								
Db	5179	AAGTATTCATTTTGTATTTTGTGAAGATTTTGAAGATTTTGAAGATTTTGAAGATTTT	5238								
QY	1966	---CCAGGTGAGTACTGTTCTCGTATTTTGTAAATATGATCTGTTCTCTTCTTGAAG-TC	2021								

```

Db      5239 TAGTGAGGTGAGTATGTTTGGTATTTGTAATATGATTTGTTTGTGTTTGAAGTTT 5298
QY      2022 CCAGAAATCAAGGGACAAATCAGTAAT -GTTATTCAGGTCATGGGATGATGGGAGTAG 2080
Db      5299 TTAGAATTATAGGGGATAATAGTATGTTTATTTAGGTTATGGGATGATGGGAGTAG 5358
QY      2081 GGCCTGAGTATTCAGAAAGTG-AACTGAGTCTCTTGATATGAATCCTTCATTACTTAG 2139
Db      5359 GGTGTGAGTATTTAGAAAAGTGAATAGTCTGTTTGATATGAATTTTATTTATTTAG 5418
QY      2140 GAAGATAACAGGCATCTCTTA-TTCACCAACTGAGGAGTGAACAGAGAAA 2191
Db      5419 GAAGATAATAGGTATTTTATTTATTTATTAATTAATGAGGATTGAATAGAGAAA 5471

RESULT 7
AB045361 148081 bp DNA linear PRI 16-MAY-2001
LOCUS Homo sapiens genomic DNA, chromosome 1q22-q23, clone:713111,
DEFINITION complete sequence.
ACCESSION AB045361
VERSION AB045361.1 GI:8918545
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Shiina,T., Ando,A., Suto,Y., Kasai,F., Shigenari,A., Takishima,N.,
Kikkawa,E., Iwata,K., Kuwano,Y., Kitamura,Y., Matsuzawa,Y.,
Sano,K., Nogami,M., Kawata,H., Li,S., Fukuzumi,Y., Yamazaki,M.,
Tashiro,H., Tamiya,G., Kohda,A., Okumura,K., Ikemura,T., Soeda,E.,
Mizuki,N., Kimura,M., Bahram,S. and Inoko,H.
Genomic anatomy of a premier major histocompatibility complex
paralogous region on chromosome 1q21-q22
Genome Res. 11 (5), 789-802 (2001)
21235190
11337475
2 (bases 1 to 148081)
Shiina,T.
Direct Submission
Submitted (30-JUN-2000) Takashi Shiina, Tokai University School of
Medicine, Molecular Life Science 2; Bohseidai, Isehara, Kanagawa
259-1193, Japan (E-mail:tshiina@is.icc.u-tokai.ac.jp,
Tel:81-463-93-1121(ex.2582), Fax:81-463-94-8884)
FEATURES
Location/Qualifiers
1. 148081
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosomes="1"
/map="1q22-q23"
/clone="713111"
/clone_lib="PAC RPC14 and 5"

ORIGIN
Query Match 12.5%; Score 274.8; DB 9; Length 148081;
Best Local Similarity 66.8%; Pred.No.9e-46;
Matches 619; Conservative 0; Mismatches 257; Indels 50; Gaps 14;

QY 531 ACGAACACCCAGAGATGTTTATGCCATATTCAGATCCTCTTTGCTTTTCTGACTTA 590
Db 102187 ATGAACACCCACATGTTGTAGTGGCTGCTATTCTAGATCCTCTGTTTCTGATT 102246
QY 591 CTCCTTTTGTAAAGTATTTCATATGTTTCAATCGAATTCAGCTATTCATCTCA 650
Db 102247 CCTGCTTTAGTGGGTTTAGACA---GTTTCCAAATCCAACTTTA---ACTCTCTCA 102300
QY 651 TGSCCCTAGTTTGGATTAAACAATCAGCTTTCTTTGCTCTGATGCGCTTGACAG 710
Db 102301 TGCTGCTGGTTGGATTAGACAATCAGCTTTATTTCTCTCTAGATGCGCTTGATCA 102360

```

```

QY 711 T-----ACATATTACAGTCTATGG-GCCCTTTCCATCATGACCATCCTCTTGTCAGA 765
Db 102361 TTTTCCCTTATTAGCTGTTTGGGATCCTTTTCCAGTCAAGACCATTTTGGCTTAAGA 102420
QY 766 TCTGCTGCTGGGTTAAACACACATCATCTGCTAGTTGATGCCATTTTATTGTGACATCTGG 825
Db 102421 TCAGCTAGTTGGTTAAACACACAGCACGTCTGCTAACTGATACCATTCATTGACAAATCAG 102480
QY 826 GGTTCACAGCAGCTAAAGGCGACCCAGTGTCTTTAACAGAGAAAATATACACTGCTGAGCA 885
Db 102481 AGTTATGATACTTTGTAGTGCACCCACTGTTCTTCACAG-GAAAAAAAATTTGCAGTACA 102539
QY 886 GAGGATTAAACAGACTCCTTCACACTACCGGAAAGTTTCAGTGCATTTTCCAGGATTTTCAAG 945
Db 102540 AAGGGTTTATCAGGCCCCCTCATCTTCAGAGTTTTCAGTGC-TGTTCCAGAAATGTTCAAG 102598
QY 946 TCTTCGATCTGCTCATGTTTGAGAGTCCACTATCCAGTTTGTGTTTATCCAGGAAATATGCA 1005
Db 102599 GTTTCAGTCTATGATGAACAA-----CTATCCATAAATAATATGCA 102639
QY 1006 GAAATCTGAGCATTTGCTGAATCTAATCATTCAGGTAAAGTAACCTAG---TCATGCTCA 1062
Db 102640 AAAAGCATGAGTAATCATGAATTTAATAATCAAGGTAAGGCGAGCTCAGTTATTATATACTCA 102699
QY 1063 AGAGGACAGCTAGTGTAAACTCCTTCATATAAATTTAAGGAAGGATTTTACATATTATA 1122
Db 102700 AGAGGACCACTAGTATGAACCTCTTTCAAAATTTATTTGAAT-AAGGATTTTACATATTATA 102758
QY 1123 ATGACTTCTTCATCTGG--AAAGCCCGAGCTTGTCTAGTTTATTAATTAATTTGGAATACT 1180
Db 102759 ATGACTTCTTCGCTCTGTCAAGCTCCAGGTGTATCAGTTTATTAATACTGGTGAATA-T 102817
QY 1181 TCAGGTAGTATTTAAACAATAGCGATTTTCATAGAAGACATTTCCCAACCCAGGAGATGC 1240
Db 102818 GTAGGTACTATTTTAAACAAA-AACAATTTTCACAGTAGACCCACCCCTATCCAGAGATGA 102876
QY 1241 AGTAAAGTATTTGGTAGCAATAGATGCTCTTACCAAGATCTCTCTGCCACTGCCACTCTCT 1300
Db 102877 AGTCAAGTTCTGGTAGCAATAGACATCTCTACCAAGAT---CTCCGTCTCTGCCACTCT 102933
QY 1301 GTGCACAGAAAGGAGGGGAGAA-----GAAGCCATGTTGCTTCTCTATATTATTAGC 1353
Db 102934 GTCAACAGAAAGGAGGGGAGAACTTTGCTGTTTTCCTTTTCCCAATTTATATTAAATC 102993
QY 1354 ATCAATTAGGGAATATAGTCTCTTAATGACATCAGATGAAGTCTGCTGGTAAGAACT 1413
Db 102994 CTATTCAAGCGGCTACCGGTTTCTAATGACCTATATGAAGTTAGCTTTTATAAGAAAGT 103053
QY 1414 TCTAATGTTGGAACCTCTTTTCAGCTA 1439
Db 103054 TCTGAGTTTGAACCTCTTTTCAGCAA 103079

RESULT 8
AC010814/c
LOCUS Homo sapiens clone RP11-3F3, WORKING DRAFT SEQUENCE, 13 unordered
DEFINITION pieces.
ACCESSION AC010814
VERSION AC010814.6 GI:8516051
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 179536)
Birken,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone Rp11-3F3
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 179536)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Balwin,J., Barna,N., Beckerly,R., Boguslavsky,I., Boukhgalter,B.,
Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,

```


QY 1123 ATGACATTTCTGATCTGG--AAAGCCAGGCTTCTGAGTTAATTAATTAATGATGATACT 1180
Db 136524 ATGACATTTCTGCTCTCTCAAGCTCCAGGTGTATCAGTTAATTAATTAATGATGATACT 136582
QY 1181 TCAGGTAGTATTAAACAATAGGATTTTCATAGAGACATTTCCCAACAGGAGATGC 1240
Db 136583 GTAGGTACTATTAAACAAA-AAACAATTTCAAGTAGACCCACCCCTATCCAGAGATGA 136641
QY 1241 AGTAAAGTATTGGPAGCAATAGATGCTCTACCAAGATCTCTCTGCTGCCACTGCCACTCT 1300
Db 136642 AGTCAAGTCTTGAGCAATAGACATCTTACCAAGAT--CTCTGCTCTGCCACTCT 136698
QY 1301 GTGCACAGGAAGAGGGAGAA-----GAAGCCAGTGTCTCTCTATATTGAGC 1353
Db 136699 GTGAACAGGAAGAGGGAGAACTTTTGGCTTTTCCCAATTTATTAATTAAGATC 136758
QY 1354 ATCAATTAAGGAATATAAGTCTCTAATGACTCATGTAAGTCTGCTGGGTAGAAAGT 1413
Db 136759 CTTATTCAAGCGGTACCGGTTTCTAATGACCTATATGTAAGTATGCTTTATTAAGAAAGT 136818
QY 1414 TCTAATGTTGAACTCTCTTTTCTGAGCTA 1439
Db 136819 TCTGAGTTGAACTCTCTTTTCTGAGCA 136844

RESULT 10
AX188367
LOCUS AX188367 2775 bp DNA linear PAT 06-AUG-2001
DEFINITION Sequence 4062 from Patent WO0142467.
ACCESSION AX188367
VERSION AX188367.1 GI:15139840
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
Schlegel, R., Deeds, J., Berger, A. and Zhao, X.
Genes, compositions, kits, and methods for identification,
assessment, prevention, and therapy of cervical cancer
Patent: WO 0142467-A 4062 14-JUN-2001;
Millennium Predictive Medicine, Inc. (US)

FEATURES
source
1..2775
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 11.0%; Score 240.2; DB 6; Length 2775;
Best Local Similarity 93.3%; Pred. No. 2.2e-38;
Matches 251; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
QY 1701 CCACATTTGTTCTTCTACTCCATTTCTCTGGGAATAGCAGATAGCAGCAGCACT 1760
Db 10 CCGCGTCTGACTAGCTGCGTGGAGAGCCCGGGAATAGCAGATAGCAGCAGCACT 69
QY 1761 AGTCAGCTAACTAAGTACTCAACAGGCTTTTCTCTGTTATCTTTGAGATACTT 1820
Db 70 AGTCAGCTAACTAAGTACTCAACAGGCTTTTCTCTGTTATCTTTGAGATACTT 129
QY 1821 CATTTTCTTAGCGTTCTGAGATTAACAATCTCTGCGTTCGGTTCTGGAACTTTAC 1880
Db 130 CATTTTCTTAGCGTTCTGAGATTAACAATCTCTGCGTTCGGTTCTGGAACTTTAC 189
QY 1881 TGATTTATCTCCCCCTCACAATAAGCATTTGTTCTGCTGATTTCTGAAGATCTCAAG 1940
Db 190 TGATTTATCTCCCCCTCACAATAAGCATTTGTTCTGCTGATTTCTGAAGATCTCAAG 249
QY 1941 ATCTGACTACTGTTGAAAAAATTTCCAG 1969
Db 250 ATCTGACTACTGTTGAAAAAATTTCCAG 278

RESULT 11
HUMIF116A
LOCUS HUMIF116A 2709 bp mRNA linear PRI 01-JAN-1995
DEFINITION Human interferon-gamma induced protein (IFI 16) gene, complete cds.
ACCESSION M63838
VERSION M63838.1 GI:184568
KEYWORDS interferon-gamma inducible protein.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (sites)
Trapani, J.A., Browne, K.A., Dawson, M.J., Ramsay, R.G., Eddy, R.L.,
Snow, R.B., White, P.C. and Dupont, B.
A novel gene constitutively expressed in human lymphoid cells is
inducible with interferon-gamma in myeloid cells
Immunogenetics 36 (6), 369-376 (1992)

JOURNAL
MEDLINE
PUBMED
COMMENT
FEATURES
source
1..2709
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="CTL/NK cell"
1..2709
/gene="IFI 16"
1..264
/gene="IFI 16"
265..2454
/gene="IFI 16"
/codon_start=1
/product="interferon-gamma induced protein"
/protein_id="AAA58683.1"
/db_xref="GI:184569"
/translation="MGKKYKNIVLLKGVINDYHFMVKSLLSNDLKMLKREYD
KIQADIMESKFGDAGLKLKIFDPTLEDLAETLKKELKVGKPALSRKKKEV
HATSPASTSTVKTGEGATGPAQKQKSTKEKAGKPKGSKVSEETQPPSPAGAMS
TAMGRSPSTJUSAPENSSSTENPKTVAKQVTPRNVLQKRPVIVKVLSTWPKPEY
ETPEMEKIMFHATVATOTFFHVKVLNTSLKEKFNKKIILISDYLEVLSLEVNEE
STVSEAGNQTFFVFNKIINRAKETLIDILHQAQSNIVGVFMLHKTNNQKTTI
EIODRGKDVVGTGQCHNIPCEBGDKLQFCFLRKNQMSKLSISMHSIFQIKKKT
NPRNDPKMKILPOEQQLPVPEASTTFPESHLRTQMPPTPSSFFTKSDTTS
KXNDFPMQILKEGSHFPPTMTSIGAESHPHTQMPSTPSSFTLLKPLKLTBP
EVSIEDSAQSDLEKSHVNLNATESFVYEPKEQKMFATVATENEVPR/KVFNIDLE
KTFPKIATANYVCRNGLEVFPFLLVADNVADRNMETPKGLIRASV7PKINQLCS
QTGSGFVNGVFEVHKKNVRGEFTYIEIQDNTGKMEVVVHGRLNTINCESGDKLTSF
ELAPKSGNTGELRSVHSHIKVITRKNKKDILNPDSSMETSPTDFP"

polyA_signal
2677..2686
/gene="IFI 16"

ORIGIN
Query Match 11.0%; Score 240; DB 9; Length 2709;
Best Local Similarity 100.0%; Pred. No. 2.4e-38;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1730 GGAATAGCAGAAATAGGAGCAGCCAGCAGTACTAGTCAGTCACTCAACCAAGG 1789
Db 1 GGAATAGCAGAAATAGGAGCAGCCAGCAGTACTAGTCAGTCACTCAACCAAGG 60
QY 1790 CTTTTTTTCTTGTATCTTTTCAGATACCTTCATTTCTTAGCGTTTCTGGAGATTCAA 1849
Db 61 CTTTTTTTCTTGTATCTTTTCAGATACCTTCATTTCTTAGCGTTTCTGGAGATTCAA 120
QY 1850 CATCTCGCGTTCCTGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAATAAG 1909
Db 121 CATCTCGCGTTCCTGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAATAAG 180
QY 1910 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGAGTACTGTGAAAAAATTTCCAG 1969
Db 181 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGAGTACTGTGAAAAAATTTCCAG 240


```
|||||
Db 121 CATCTGGGTTCCGTTTCTGGAACTTACTGATTTATCTCCCCCTCACAAATAAG 180
QY 1910 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCAG 1969
Db 181 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCAG 240

RESULT 14
AX884965
LOCUS AX884965 504 bp DNA linear PAT 18-DEC-2003
DEFINITION Sequence 828 from Patent EP1033401.
ACCESSION AX884965
VERSION AX884965.1 GI:40040653
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 Dumas Milne Edwards, J.B., Duclert, A. and Giordano, J.Y.
Expressed sequence tags and encoded human proteins
JOURNAL Patent: EP 1033401-A 828 06-SEP-2000;
Genset (FR)
FEATURES
source Location/Qualifiers
1..504
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
264..>503
/note="unnamed protein product"
/codon_start=1
/protein_id="CAE99079.1"
/db_xref="GI:40040653"
/translation="MSYVMGKKYKNIVLLKGLVINDYHFRMVKSLLSNDLKLKMR
EYDKIQIADIMEEKPRGDAGLKLKIFEDIPLE"
ORIGIN
Query Match 10.9%; Score 239.4; DB 6; Length 504;
Best Local Similarity 99.6%; Pred. No. 4.7e-38;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1729 GGGGAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 1788
Db 11 GGGCAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 70
QY 1789 GCCTTTTTCCTTTTCTTGTATCTTTCAGATACCTTCAATTTCTTTCAGGATTTCTGGAGATTACA 1848
Db 71 GCCTTTTTCCTTTTCTTGTATCTTTCAGATACCTTCAATTTCTTTCAGGATTTCTGGAGATTACA 130
QY 1849 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 1908
Db 131 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 190
QY 1909 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCA 1968
Db 191 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCA 250
QY 1969 G 1969
Db 251 G 251

Search completed: March 28, 2004, 05:37:50
Job time : 8599 secs

|||||
Db 121 CATCTGGGTTCCGTTTCTGGAACTTACTGATTTATCTCCCCCTCACAAATAAG 180
QY 1910 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCAG 1969
Db 181 CATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCAG 240

RESULT 15
BD024575
LOCUS BD024575 504 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD024575
VERSION BD024575.1 GI:22565798
KEYWORDS JP 2001269182-A/821.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE
1 (bases 1 to 504)
AUTHORS Edwards, J.B.D.M., Duclert, E. and Jordan, J.Y.
TITLE Sequence tag and encoded human protein
JOURNAL Patent: JP 2001269182-A 821 02-OCT-2001;
GENSET
COMMENT
OS Homo sapiens (human)
PN JP 2001269182-A/821
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLERT, JEAN YVES
PI JORDAN
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC Key Location/Qualifiers
FH CDS 264..503.
FT source Location/Qualifiers
1..504
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 10.9%; Score 239.4; DB 6; Length 504;
Best Local Similarity 99.6%; Pred. No. 4.7e-38;
Matches 240; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1729 GGGGAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 1788
Db 11 GGGCAATAGCAGATAGGAGCAAGCCAGCACTAGTCACTAACTAAGTCACTCAACCAAG 70
QY 1789 GCCTTTTTCCTTTTCTTGTATCTTTCAGATACCTTCAATTTCTTTCAGGATTTCTGGAGATTACA 1848
Db 71 GCCTTTTTCCTTTTCTTGTATCTTTCAGATACCTTCAATTTCTTTCAGGATTTCTGGAGATTACA 130
QY 1849 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 1908
Db 131 ACATCCTGCGGTTCCGTTTCTGGAACTTTACTGATTTATCTCCCCCTCACAAATAA 190
QY 1909 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCA 1968
Db 191 GCATTGATTCCTGCATTTCTGAAGATCTCAAGATCTGGAATCTGTGAAAAAATTTCCA 250
QY 1969 G 1969
Db 251 G 251

Search completed: March 28, 2004, 05:37:50
Job time : 8599 secs
```